

The Case for Moral Objectivism

A Moorean Odyssey through Metaethics



Ainar Petersen Miyata

Thesis presented for the degree of
MASTER OF PHILOSOPHY

Supervised by
Associate Professor Sebastian Watzl

Department of Philosophy, Classics, History of Art and Ideas
Faculty of Humanities

UNIVERSITY OF OSLO
Spring 2015

Objectivity is the central problem of ethics.

—Thomas Nagel, *The View from Nowhere*

Here is one hand. And here is another.

—G. E. Moore, *Proof of an External World*

You may think that you're a moral relativist or subjectivist,
but I don't think you are.

—David Enoch, *Why I am an Objectivist about Ethics*

How could all these people be so mistaken?

—Derek Parfit, *On What Matters, Vol. 1*

Ainar Petersen Miyata

The Case for Moral Objectivism

A Moorean Odyssey through Metaethics

Abstract

One of the central questions in metaethics is whether morality is a fundamentally subjective or fundamentally objective domain. Are we, when we are discussing moral claims or making moral judgments, engaging in an activity where there are standards of correctness that are independent of our own subjective beliefs, wishes and desires? In this thesis, I will attempt to answer this question using an argumentative strategy borrowed from G. E. Moore: his famous proof of an external world by the waving of his hands.

While Moore's argument might seem simplistic, the philosophical strategy behind it is deeply interesting. Moore took a claim from within the domain that was being doubted, and used it in an argument against a metatheory about the domain. This what I will do in the case of the moral domain. I will take premises from within morality itself, and use them in arguments against the metaethical theory of subjectivism. The specific arguments that result will be very simple. The discussion of what the use of such arguments can show will be quite complex.

In the end the strategy of employing Moorean arguments against subjectivism will result in two distinct reasons to reject the theory, one from the independent force of the arguments themselves, and one from what the possibility of formulating such arguments against subjectivism says about the theory. Together, these reasons give us a relatively strong reason to reject subjectivism, and to embrace objectivism about morality.

© Ainar Petersen Miyata 2015

The Case for Moral Objectivism: A Moorean Odyssey through Metaethics

<https://www.duo.uio.no/>

Acknowledgements

It is hard to believe that I finally am in the position where I can give thanks where thanks are due, and supporting the performatives by engraving them into a finished product. Or print them on it. Or whatever. I am finished; I'm allowed to say what I want. I have a feeling that I would never have been able to get to where the thoughts turned to matter, and matter into a fully grown thesis, if it wasn't for all the excellent people I traveled with, and had help from, on the way.

First, I must thank all of the members in the colloquium of metaethical aspirants I was fortunate enough to be a part of. The often lively discussions we had—every other week for more than one and a half year by now—have unquestionably been one of the greatest influences on my thoughts and this thesis. Conrad Bakka, Martin Ravneberg, Marcus Bøhn, Øyvind Sætre Strøm, and, occasionally, Caj Strandberg: you are the chorus that whispers between the pages of this thesis. Reader, don't listen to them—they all think I'm that crazy objectivist guy.

Then there are friends and family, so many and so diverse, often ready to listen to me talk in incomprehensible tongues about my fascination with waving hands, and something called normativity? What is that again—a disease? My uncles, aunts, and cousins, my two wonderful sisters, my brother-in-law and, of course, my mom and dad. (Mom, I hope you'll get through the whole thing, and dad, I hope you at least get this far. You know, I always want to make you proud.) Thanks to all the Vedvik-brothers and Sivert for distracting me with thoughts of God, thanks to Christine for all those lovely cups of tea that always ended up cold. And thanks especially to Daniel Gitlesen and Thomas Rambø for reading and proofing my last drafts.

Thank you, Mariona and Frøya, with all my love and all my heart: you are the most wonderful beings I have ever known.

And finally, Sebastian Watzl, without whose constant positivity, encouragement and criticism the day of writing this acknowledgement might never have come to pass: You made me feel like a philosopher. For that, I owe you a beer. At the very least.

A.P.M.

Oslo,

June 9, 2015

Table of Contents

Abstract	4
Acknowledgements	5
1. Objectivity and Subjectivity	8
1.1 What is Metaethics?	8
1.2 What is Objective and what is Subjective?	10
1.3 Why Objectivism?	17
1.4 Some Objectivist Theories of Metaethics	20
1.5 Why Subjectivism?	23
1.6 Summary and Preview	32
2. Moorean Arguments Against Subjectivism	33
2.1 Basic Moorean Arguments	34
2.2 Revisionary Theses and the Limits of Moorean Arguments	37
2.3 Moorean Arguments Against Metaethical Subjectivism	41
2.3.1 Moorean Arguments Against Simple Subjectivism	42
2.3.2 Improving Subjectivism: Group-Relativity	45
2.3.3 Improving Subjectivism: Deliberation	48
2.4 Summary and Preview	50
3. The Best Subjectivist Theories	51
3.1 What Is Idealized Deliberative Subjectivism	52
3.2 Williams-Style Subjectivism	53
3.3 A Moorean Argument From a Moral Premise	58
3.3.1 Arguing Against the Linking Premise	59
3.4 Railton-Style Subjectivism	61

3.4.1 The Schema for Modifying S.....	63
3.4.2 A Moorean Argument From a Normative Premise.....	65
3.4.3 The Sanity Condition	67
4. Evaluating the Arguments.....	69
4.1 The Case Against Tolerable Subjectivism.....	70
4.1.1 The Inductive Case against Subjectivism	71
4.1.2 Are the Subjectivist Theories Radically Revisionary?	74
4.1.3 The Argument Against Tolerable Subjectivism.....	78
4.1.4 Should We Reject Radically Revisionary Theories of Morality and Normativity?	78
4.1.5 A Response from the Subjectivist.....	81
4.2 How Strong Are the Moorean Arguments?	82
4.2.1 An Age of “Post-Moorean Modesty”?.....	82
4.2. Two What Makes a Fact “Moorean”?	83
4.2.3 The Debunking Objection.....	88
4.2.4 Improving the Debunking Argument.....	90
4.2.5 Responding to the Debunking Argument	93
4.2.6 The Status of the Moorean Arguments against Subjectivism.....	96
4.3 Summary and Preview	97
5. Subjectivism, Rejected.....	98
5.1 In Favor of Objectivism	98
5.2 In Favor of Subjectivism.....	99
5.3 Against Subjectivism	99
5.4 Conclusion	100
Bibliography	101

1. Objectivity and Subjectivity

Objectivity and subjectivity are, like most interesting concepts in philosophy, intuitively compelling but hard to conclusively pin down. As already mentioned, there are at least a few distinct ways of interpreting these notions,¹ and if we are trying to be philosophically precise it is probably a bad idea to try to capture all these different senses at once. To do real philosophical work we will do better if we choose one or two of them to focus on. In large part because of its multiple senses, the objective/subjective-distinction may be one of the least clear among the many distinctions and terminological pairs that carve up the landscape of metaethics. Despite this, it seems to me that the distinction reflects an idea that is tremendously important to the way we should think about ethics. Moreover, the distinction appears often in the contemporary literature, and is worth discussing for this reason alone. I shall try to choose carefully a definition that both makes sense of the use of the distinction in the literature and captures the idea that makes the distinction interesting in the first place. First, however I have to say a few words about what I mean by metaethics.

1.1 What is Metaethics?

I see metaethics as the field of philosophy that studies what it is that we do when we do ethical theory or engage in moral discourse. The relationship is somewhat analogous to that between philosophy of science and science itself. Science, in this context is what we might call a first-order discipline.² It is concerned with building or discovering theories to describe the world in scientific terms and, crucially, in discovering the *right* theories. It is also concerned with applying those theories and scientific principles to concrete situations and problems in the world. Some people would make a distinction here, between, say, science and technology, reserving “science” for the former activity. Correspondingly, ethics is a first-order discipline that is concerned with building/discovering the right *ethical* theories to describe the world in moral terms, and also with solving concrete ethical problems, such as whether or not euthanasia is

¹ Simon Kirchin lists four (Kirchin, 2012: 25–26, 132).

² See (Miller 2013:1–3) for a similar, more detailed account.

immoral. Here too, some people like to make a distinction between ethical theory and applied ethics, reserving the term “ethical theory” for the first type of activity.

Philosophy of science is a second-order discipline that is concerned with what we really do when we do science. Some questions asked in philosophy of science are: “are the entities that scientists talk about, such as electrons, really real?” and “what are we really doing when we say that one scientific theory is better than another?” Similarly, in metaethics we ask such questions as “what are we really doing when we say things like ‘murder is wrong’ or ‘Kantian deontology is the one correct moral theory’” and “do moral facts exist, and if so, what kind of things are they?”

It should be obvious that the main question of this thesis, “is the moral domain objective or subjective?”, falls under the heading of metaethics on this picture. I am not here doing work on ethical theory or attempting to solve a practical ethical problem, but rather asking about what the subject matter of the first-order domain of ethics is like. Are there objective standards of correctness for claims such as “murder is wrong” and “Kantian deontology is the one correct moral theory”? This is not normally a question asked comfortably within the domain of first-order ethics.

When it is laid out like this, it might perhaps seem that there is a sharp divide between first- and second-order ethics, with applied ethics and ethical theory on one side and metaethics on the other. Some philosophers think that there is such a divide, or at least that there should be. The thesis that there is, or should be, such a divide can be called the moral neutrality of metaethics.³ I, however, think that the lines between metaethics, ethical theory and applied ethics are well and thoroughly blurred. I think that anyone in the business of doing ethical theory or applied ethics should be intently concerned with the answer to at least certain metaethical questions. If it turns out that there are no objective standards of correctness for claims such as “Kantian deontology is the one correct moral theory,” then it seems to me that we have a good reason to reject the claim outright. If, say, the standards of correctness for this claim are subjective, then the truth of the claim could vary depending on who you ask. But surely, this is contrary to the very meaning of “Kantian deontology is the one correct moral theory”! A better analysis of the situation would be to say that much, or most, of our first order moral discourse carry certain metaethical commitments. Accepting theories like Kantian deontology or

³ For an early discussion of the moral neutrality of metaethics, see Gewirth (1968).

utilitarianism usually carry objectivist implications. And, if the main arguments in this thesis are successful, it should be clear that the reverse is also true—metaethical theories carry first-order commitments.⁴

Although I will mostly be concerned with metaethics in this thesis, it will become clear that the questions and arguments I am discussing can also be thought of as being concerned with *metanormativity*. I will understand metaethics to be a sub-field of metanormativity, so all metaethical questions are necessarily also metanormative questions. The metanormative domain includes all things that have to do with normative reasons and value, where the metaethical domain only includes those that have to do with moral reasons and moral value. When I talk about normativity rather than morality in this thesis, I will generally be talking about what we might label *prudential* normativity, which is the sub-domain of normativity concerned with what we might describe as “what is in the best interest of agents.” Where I talk about “normativity in general,” I mean the whole domain of normativity, not restricted to morality or prudential normativity. I will make it clear whether I am talking about just morality, prudential normativity or normativity in general when this is important to the discussion, but often, I could easily have replaced “morality” with “normativity in general” without impacting the discussion overmuch.

1.2 What is Objective and what is Subjective?

To prime our intuitions about what objective and subjective domains are like, I will begin this section by presenting a few examples of domains that are paradigmatically objective or subjective. But first a few clarifications.

By “domain” I simply mean a roughly recognizable area of discourse, and it is not supposed to be any sort of metaphysically significant notion. It is sometimes the case that the borders between domains correspond to metaphysically important divides, for instance the divide between the domain of the normative and the domain of the non-normative, but this is not required. Basically, any area of discourse that it would be useful to talk about under a unified heading we can call a domain. So, if I were the owner of a chocolate-factory, I might talk about the domain of candy, its sub-domain of chocolate and *its* sub-domain of coffee-flavored

⁴ David Enoch also concludes that the moral neutrality of metaethics is false based on the same type of reasons in his (2011).

chocolate. Another useful notion in this regard will be that of a *claim*, by which I simply mean a unit of discourse, either uttered or unuttered, which you might say is something that expresses a proposition (but you don't need to, in case you don't believe in propositions). We can identify claims as belonging to domains and, since a domain really is not much more than the set of all possible claims that belong to it, it is the properties of these claims that give us the characteristics of the domain. All the claims belonging to domain must share at least one property that allow them to be identified as belonging to the domain. For example, all claims belonging to the chocolate domain, i.e. chocolate-claims, must be about chocolate in some way. (Again, since domains do not necessarily reflect metaphysically interesting quantities, there need not be any *interesting* unifying properties of their claims).

With that out of the way, let us move on to a couple examples of objective domains:

The natural sciences: This is a domain that few would hesitate to call objective. When we evaluate scientific claims, we expect them to conform to all sorts of objective standards, and we do not accept scientific theories just because we like them: they have to at the least exemplify some virtues that seem to indicate objective standards of correctness, like predictive and explanatory power.⁵ Take quantum mechanics. If all the meticulous calculations done in the field resulted in were some numbers that the scientists found esthetically pleasing, then no one would take the field seriously as a theory of physics. Quantum mechanics yield impressively accurate predictive results, allowing us to do things like predict (probabilistically) the behavior of subatomic particles and build computers that can perform calculations through quantum superpositions.

To take a simpler example, it seems clear that there are objective standards of correctness for deciding the appropriate melting point of a particular element on the periodic table. Even if the entire scientific community decided on making the melting point of iron 0° C, this would not make iron melt at the same temperature as water. In other words, *there is something about the world that imposes constraints on what successful science is, something more than what a person, or a group of people prefer, like or want.*

Mathematics: Mathematics is an objective domain. $2 + 2 = 4$ even if you don't like that it does. Pythagoras' theorem is accepted as a theorem, not because mathematicians like it, but because it is provable. It also yields the right results when applied correctly, and it can be used to predict how things are in the real world. Just like with science, *there is something about the world that imposes constraints*

⁵ Although, as I am sure many philosophers of science would like to point out, it is by no means easy to spell out exactly how or why any of the criteria for being a good scientific theory amounts to more than saying: "It is good because scientists like theories like this."

on what successful mathematics is, something more than what a person, or a group of people, prefer, like or want.

Let us now look at an example of a subjective domain:

Favorite things: This is, I believe, the most clearly subjective domain that there is. There are many subdomains within the domain of favorite things, such as those of favorite foods, favorite music and favorite colors. For example: I like red, black and deep purple. I also like the green of spruce woods in foggy spring. You might like different colors. My claims about which colors are my favorites are not subject to any standards of correctness other than the facts about what colors I actually *like* best. I might present some story about how I came to like that particular shade of green, but *that story is in an important sense no better than a simple "I just do" as a justification of my claim.* The only standard of correctness that applies to claims within the domain of taste is how the claim fits with the things that certain people like. The way we evaluate whether or not a claim like "red is my favorite color" is true is by figuring out whether or not the utterer likes red better than all the other colors. If she does, then it is true, if she doesn't then it isn't.

Let me forestall one possible objection. You might say: "but, we can ask things like 'is red the prettiest color?' or 'does ice cream taste better than dirt?', and these things look like they could have objective answers." This is indeed true, and it is why I limited the domain to that of favorite things, rather than choosing taste or aesthetics in general. Whether something is one of your favorite things or not, does really just depend on whether or not it is one of the things that you like best.

A domain can either be more like that of the natural sciences and mathematics, or it can be more like the domain of taste. Hopefully, these examples have will leave the reader with the sense that there is a salient distinction between objective and subjective domains, and that there are some domains that quite clearly are objective and likewise some, or at least one, that is clearly subjective. We can now frame the question I will try to answer in this thesis thus: Is morality more like the paradigmatically objective or subjective domains? In order to get more philosophical work done, however, a precise definition of objective and subjective domains is desirable. Following the clues left by the presentation of the examples above I will now attempt to provide this, starting with the definition of a subjective domain.

What makes a domain, like that of favorite colors, subjective? Take a claim within the domain of favorite colors, like "red is my favorite color." We can now ask what exactly must be the case for this claim to be true, or, put a bit differently, what kind of facts must obtain in order for this claim to be correct. What we are looking for here are the *standards of correctness* for the claim. For a claim like "red is my favorite color," the standards of correctness are simply

whether or not I like the color red better than any other color, or something similar. Accordingly, the facts that must obtain are some facts about what I like. These are a type of facts that we could call *subjective facts*. I define subjective facts as follows:

Subjective facts: Facts about what a person, or a group of persons, thinks, wants, likes, desires, prefers, aims to do, etc.⁶

We can see that all the standards of correctness of claims that belong to the domain of favorite colors refer to subjective facts about actual agents.⁷ And now we already have what we need for our definition of a subjective domain:

Subjective domain: A domain is subjective just in case all of the standards of correctness for claims belonging to the domain refer to subjective facts of actual agents.⁸

Notice that the definition is quite strict, implying that, if there is even one claim belonging to that domain that has standards of correctness that do not refer to subjective facts, the domain is not completely subjective. This, however, is fine, since it is completely reasonable to suppose that domains can be subjective to different degrees, and that it is not all that common for a domain to be completely subjective. We can thus talk about partly or mostly subjective domains, for

⁶ We could also call these facts *attitudinal facts*, since they are all facts about what attitudes (in the philosophical sense) people or groups of people have.

⁷ The reference to actual agents is crucial, but the reason for this will not become entirely clear until we start to discuss subjectivist theories about metaethics. But we can get an idea of it. Claims like “red is my favorite color,” is subjective precisely because whether it is true depends on what colors I actually like. If the actuality-requirement was removed, we could do this: we define *favorite color* as those colors that, if I lived in a world where everyone liked red the best, would be the color I liked best. The standard of correctness here also refers to a subjective fact, namely a fact about what I would like best if I lived in a world where everyone liked red best, and so the claim would still be count as subjective. But on this weird definition of favorite colors, it just seems weird to say that what my favorite color is, is subjective—it really isn’t. Whatever I think or whatever colors I like just don’t enter into the equation, and red will be my favorite color no matter what. For that matter, your favorite color would also be red and everyone else’s too. Such a view about favorite colors would, of course, just be plain weird, and part of the reason for its weirdness is that it would be an objectivist account.

⁸ There is a small complication. Take a slightly unusual claim that we nevertheless might want to say belongs to the domain of favorite things: “A’s favorite color is the color that is spelled with three letters and starts with the letter R.” If it is the case that A likes the color red best of all, then this claim will only be true if certain facts about the English language are true, such as what three-letter color words exist that start with the letter R, and these are obviously not subjective facts. However, since the claim then would simply be a roundabout way of saying “A’s favorite color is red,” I don’t think the case presents a real problem. After all, “A’s favorite color is red” is only true if certain facts about what the word “red” means is true, and so it looks like facts that fixes the referents of elements in the claim should not be a part of the standards of correctness that we are looking for.

instance, where those are domains where respectively some and most standards of correctness refer to subjective facts.

An objective domain is then a domain where all of the standards of correctness for its claims refer to objective facts, where objective facts just are facts that are not subjective facts.

Objective domain: A domain is objective Just in case all of the standards of correctness for claims belonging to the domain refer to objective facts.

Again, the requirement for being a completely objective domain is very strict, and, again, this is just what we want. As with subjective domains, we have different degrees of objectivity in domains, where a domain that is more weakly objective also has some standards of correctness that are subjective. These definitions classify the domains we have already discussed correctly, with the domain of favorite things being completely subjective, the domain of the natural sciences and the domain of mathematics being completely objective. There is one quirk with this definition, however, and that is that the science of psychology comes out as very subjective. This just seems wrong, since empirical psychology is no more subjective than any other empirical science.⁹ This would indeed be a huge problem if the intent was for the definition of objective and subjective domains to capture a fundamental metaphysical joint in nature. Our purpose for the definition is far from such an extravagant task. We only need it to enable us to talk precisely about certain types of metaethical theories, and, for this, it will do nicely. If you are still worried, however, think of it like this: all domains that are subjective on our definition, even that of favorite things, are objective in one sense: the standards of correctness for their claims are *themselves* objective. Although the standards of correctness *refer* to subjective facts, they are not themselves subjective, and I cannot choose whatever I want to be the standard as the standard of correctness for a subjective domain—they are objectively given. There is an objective answer to the question “is red my favorite color?” So psychology will still come out as objective in this higher-order sense, even on our current definition. And, as I indicated above, there is no reason to suppose that our definitions will capture *all* the senses of objectivity or subjectivity.

⁹ Although, as Conrad Bakka helpfully pointed out to me, psychology is often *not* placed among the so-called hard sciences. And as I say below, much of the what is meant by the hard/soft-distinction between sciences can be explained in terms of the objective/subjective-distinction. This means that my definition might have classified psychology correctly after all. This just goes to show that the objective/subjective-distinction is a tricky creature.

Using these definitions, we can formulate the metaethical and metanormative theses of subjectivism and objectivism:

Subjectivism about morality/normativity: The thesis that morality/normativity is a subjective domain.

Objectivism about morality/normativity: The thesis that morality/normativity is an objective domain.

One slightly tricky issue remains: While we could say that we only care about the “pure” forms of objectivism and subjectivism, those that claim morality is a completely objective and subjective domain respectively, this would simply exclude too much of the metaethical/metanormative landscape. Very few theories can be classified as either pure objectivism or pure subjectivism, and these theories are not necessarily the most plausible ones. Yet, if most of the interesting theories are “impure” versions, then they would classify as both subjectivism and objectivism! Granted, we still have a spectrum of theories that go from more subjectivist to more objectivist, but this does not give us the crisp distinction we need. I therefore propose the following additional category, which we can call *strong subjectivism*, which is not as strict as a pure subjectivism, because it allows claims to have some standards of correctness that refer to objective facts, but that requires each and every claim within the domain to have *at least one* standard of correctness that refers to subjective facts.¹⁰

Strong subjectivism about morality/normativity: All the claims within the domain of morality/normativity have at least one standard of correctness that refers to subjective facts about actual agents.

From this point on, I will mean strong subjectivism about morality/normativity whenever I talk about subjectivism. This classification is important because the main reasons for favoring a subjectivist theory will require that theory to be at least strongly subjectivist. Furthermore, being less than strongly subjectivist will make a theory inherit most of the problems thought to affect objectivism. From this point on, whenever I talk about subjectivist theories I will mean strongly subjectivist theories, and whenever I talk about objectivist theories I mean theories that are not

¹⁰ For example, the versions of subjectivism presented in chapter 3 will have standards of correctness that refer to a lot of facts that are not subjective facts about actual agents. But to remain subjectivism in the sense we want, they need to always also refer to subjective facts about actual agents. An example of a theory that would be *less* than strongly subjectivist, is one where the standards of correctness of *some* moral claims are completely objective, while others are subjective.

strongly subjectivist, i.e. theories where not all claims have standards of correctness that refer to subjective facts.

Before we move on, I will mention another way to cash out the same distinction. This is in terms of the notion of *metaphysical grounding*, which is a notion closely related to explanation.¹¹ This will be useful for our discussion in later chapters, and cashing out the idea of subjectivity in this way really gets at the spirit of what subjectivist theories try to embody.¹² If you think that If A grounds B, then we can say that it is *in virtue* of A obtaining that B obtains, or that A *makes it the case* that B, or that A obtaining *explains* B obtaining. So, for instance, if mind-brain identity theory is correct, the fact that a certain event happened in my brain grounds the fact that I experienced an itch on my left shoulder.¹³ A subjectivist theory about morality, then, will be one that claims that moral facts are grounded by subjective facts. Since we use subjectivist theory in the strong sense we can say that a subjectivist theory is one that claims that all moral facts are grounded by at least one subjective fact. An objectivist theory of morality will be one that claims that moral facts are grounded by objective facts. And since we are using objectivism in the sense of being the negation of strong subjectivism, we can say that an objectivist theory is one that claims that not all moral facts are grounded by subjective facts.¹⁴

It does not take much for a theory to be objectivist in the sense we just defined. However, even though the criteria for being subjectivist in the sense we use are pretty strict, it is not necessarily immediately obvious when we might be dealing with a subjectivist theory instead. Take any theory that maintains that there is a right answer to any moral question, like “is it right to kill animals for food?” One might think that any such theory would be objectivist, since if

¹¹ See (Bliss, 2014).

¹² If you don’t like the notion of grounding, you can simply think of it as a relation of explanation. Nothing substantial turns on this in the thesis.

¹³ Another example: The fact that P and the fact Q grounds the fact that P and Q. The two conjuncts grounds the conjunction. Thanks to Conrad Bakka for this suggestion.

¹⁴ The way I have defined subjectivism and objectivism now makes the distinction very close to a few other distinctions that are often talked about in metaethics, namely mind-dependence/mind-independence, attitude-dependence/attitude-independence, judgment-dependence/judgment-independence and response-dependence/response-independence. Depending a bit on whom you ask, these distinctions sometimes amount to the same thing and sometimes don’t. Of the distinctions mentioned, the response-dependence/independence and attitude-dependence/independence distinctions are probably the closest. In this text I will not be overly concerned with how these distinctions interact, nor do I think that they necessarily are any better or worse than the one I use here. I have chosen to discuss the matter in terms of the objective/subjective-distinction in part because I find it slightly more intuitively compelling, and in part because I started out reading literature that used the objective/subjective-distinction, e.g. Parfit (Parfit, 2011a, 2011b) .

there is a right answer, then this answer must be right in virtue of meeting certain criteria of “rightness.” That sounds a bit objectivist. But, of course, the theory could still be subjectivist if the only criterion for rightness is something like “if you like to do it, then it is right to do it.” So, in one sense, even on subjectivist theories, there are objective answers to questions about what is right and what is wrong, etc. But it is still appropriate to call the theory subjectivist, since any right answer to moral questions would be subject to a standard of correctness that refers to the subjective fact of what you like to do. Equally, moral facts, on such a theory, would be grounded in the subjective fact of what you like to do. So this theory would make morality a lot like the domain of favorite things, where the facts about what your favorite things are depend on what you like best. And even if there are objective answers to questions about what your favorite things are, the domain of favorite things is inescapably subjective.

1.3 Why Objectivism?

There are good intuitive and commonsensical reasons to think that morality is an objective domain. First, it just seems plain wrong to think that morality is like the domain of favorite things. If, when discussing a moral claim such as “murder is wrong,” you argue that it is a good or right claim simply because you like it, few people are going to take you seriously. Whenever we discuss moral claims, we expect there to be a serious debate, and we treat this debate very differently than we would have treated a debate whose topic was “red is my favorite color.” But this is by no means conclusive evidence in favor of objectivism: we have similar debates about things like music or styles of architecture, and whether or not these domains are objective is a controversial matter. Still, if we really sit down and think about it, I bet that most of us would think that there really is more at stake in debates about morality than about architecture.

In a fun and accessible essay, David Enoch (2014) mentions three ways we can test our intuitions about what is and is not an objective domain. The first is the “spinach-test,” and it revolves around a joke of the following format:

A child hates spinach. He then reports that he’s glad he hates spinach. To the question “Why?” he responds: “Because if I liked it, I would have eaten it; and it’s yucky!” (Enoch, 2014: 2)

The joke seems to work only because whether or not spinach will be “yucky” to him is completely dependent on the very subjective fact that he is positing as the antecedent in his counterfactual conditional: If he liked it, then it would not have been yucky! So it seems like he

is making a sort of logical mistake, which we might find amusing. If there is no such logical mistake, then the joke is not funny. But a required condition for this mistake is that “it’s yucky” is a claim whose truth depends on a subjective fact. We can then try to see if the joke works if we rephrase it to be about the domain of morality:

Suppose someone grew up in the US in the late twentieth century, and rejects any manifestation of racism as morally wrong. He then reports that he’s happy that that’s when and where he grew up, “because had I grown up in the 18th century, I would have accepted slavery and racism. And these things are wrong!” (Enoch, 2014: 2).

To Enoch’s and my own ears, this version of the joke is simply not funny. And it is not because jokes about slavery and racism just don’t tend to be funny in general. It is because there does not seem to be any sort of logical mistake involved, like in the spinach-version. Here, the antecedent of the counterfactual conditional does not seem to be something that has any bearing on whether slavery and racism is wrong. The joke only seems to work if you make it about something from a subjective domain, and it doesn’t work in the case of morality.

Enoch’s second test is one concerning the phenomenology of disagreement. Here, he asks us to consider what it feels like to disagree about claims in the domain. In cases of disagreement about subjective claims, like whether or not ice cream is better than chocolate, it feels like you are just saying something about what *you* like, and perhaps are trying to influence the one you are talking to into having preferences more like your own. It’s not really like you are trying to convince the other that ice cream is essentially superior to chocolate, completely independently of what you feel about it. But in disagreements about claims in objective domains, it feels like this is exactly what you are doing. If you are disagreeing about whether rubidium has atomic number 36 or 37, this really feels like you are disagreeing about something that has an answer that is completely independent of what you think or feel about it. It doesn’t really matter what you *or* I think. Rubidium really does have atomic number 37. And the same thing seems to be true for disagreement about moral claims. If you are disagreeing with someone about a moral issue, say, whether or not abortion is wrong, doesn’t it feel more like what you are trying to get at has an answer that is independent of what you think or feel about it? At least that is what Enoch thinks, and I share his opinion, (Enoch, 2014: 6).

The third test is a “what if” test. Consider the fact that yellow cross-gartered stockings for men are out of fashion.¹⁵ Ask yourself, “what if we all had worn yellow cross-gartered stockings

¹⁵ Alas, poor Malvolio!

and almost everyone thought that yellow cross-gartered stockings were cool, would yellow cross-gartered stockings still be out of fashion?” The answer seems to be “no.” This is because what is fashionable is a subjective domain, where facts about what is fashionable are grounded by facts about what people *think* is fashionable. Now, let us try this with a moral claim, like “it is wrong to discriminate based on gender.” We ask: “What if we all discriminated based on gender, and everyone thought that discriminating based on gender was a morally acceptable thing to do, would discriminating based on gender still be morally wrong?” To me, the answer seems to be “yes.” And this indicates that our intuitions tell us that morality is not a subjective domain (Enoch, 2014: 8).

All of this shows, I think, that without some specific arguments to the contrary, some form of objectivism is the sensible place to start when looking for a metaethical theory. Much like science, and certainly a lot like most other areas of philosophy, first-order ethics is an area of study that in many respects takes for granted that there are objective answers to the questions it pursues. It is a good bet that, if you ask people who do ethical theory or practical ethics, most will have some objectivist intuitions. They think that their arguments truly *matter*, because they think that there is a right answer to hard questions like “is euthanasia wrong?,” an answer that depends on more than whether we think that it is wrong or not. As Enoch puts it, morality aspires to objectivity (Enoch, 2014: 12).

Of course, it might just turn out that morality is a subjective domain, despite the aspirations to the contrary. As I have already said, even subjectivism as we have defined it, allows for objectivity in one sense of the word, namely that there are definite standards of correctness for claims made on subjectivism. Perhaps a good metaethical subjectivism might show us how this kind of objectivity is enough. Or, perhaps, it will turn out that morality can’t be what first-order ethicists and others who care deeply about moral questions want it to be. Perhaps the sensible thing to do is just to give up on the aspiration to objectivity, faced with the weight of evidence in favor of subjectivism. It is finding an answer to these questions that is what I hope to contribute to in this thesis.

1.4 Some Objectivist Theories of Metaethics

To give a better sense of how the objective/subjective-distinction as I have defined it divides up the metaethical landscape, I will provide some examples of metaethical theories that are objectivist on my definition. First, there is a whole family of theories that postulates that there is a distinct realm of moral facts. This type of view is sometimes called Platonism, and it is often characterized as being ontologically non-naturalist, because it requires the existence of non-naturalistic entities (i.e. *sui generis* moral facts). I will say a bit more about naturalism and non-naturalism in section 1.5 below. To this family belongs those views that are called intuitionism, defended in its traditional form by of people like G. E. Moore in *Principia Ethica* (1903), Henry Sidgwick in *The Method of Ethics* (1907) and David Ross in *The Right and the Good* (1930) among others. In recent times there has been a resurgence of interest in intuitionism, with several philosophers taking up the torch, often with book length treatments, such as Robert Audi's *The Good in the Right* (2004) and Michael Huemer's *Ethical Intuitionism* (2005). Then there are those philosophers who do not explicitly say that they are intuitionists, but who hold views that are very similar, such as David Enoch with his "Robust Metanormative Realism," defended in *Taking Morality Seriously* (2011), as well as the type of reasons primitivism defended by Thomas Scanlon in *Being Realistic about Reasons* (2014) and Derek Parfit in *On What Matters* (2011a and 2011b). All of the views above can be fairly called non-naturalist realist theories, and may also be sensibly classified as platonist. These views are quite clearly objectivist: all of them maintain that the truth of moral claims are grounded in a distinct type of objective fact, and there is really no mention of subjective facts at all. These kinds of view are also ones that are often thought to face the most serious difficulties on several levels, such as ontology, epistemology and regarding the issue of moral motivation. Being non-naturalist is also often considered a huge burden on its own, but it is not completely clear that this can be considered an additional burden rather than being a concern already covered by the issues of ontology and epistemology. More on this below.

Another type of non-naturalist realist and objectivist view is that which is sometimes called sensibility-theory, (Darwall, Gibbard and Railton, 1992: 131), advocated by John McDowell (1998) and David Wiggins (1998). Many philosophers will probably object to my labeling of this view as "objectivist," especially since it is sometimes even presented under the name "subjectivism" (see for instance the chapter on "A Sensible Subjectivism" in Wiggins

1998). However, at least on my definition of subjectivism, this is a mischaracterization. One formulation of the core of sensibility theory goes as follows:

Some action/situation/person is morally good if and only if a normal/standard/best/etc. observer would judge it to be morally good in normal/standard/best/etc. conditions. (Kirchin, 2012: 129)

Although it looks like all moral facts are going to be grounded in the responses of moral agents, on this theory, there are two parts of this particular formulation that will make the theory objectivist, namely the two references to “normal/standard/best/etc. conditions.” To be informative, these criteria have to take the form of what I call *substantive list criteria*, which I will say more about in chapter 2 and 3 (see especially section 2.3.2). All theories that employ such substantive list criteria, either explicitly or implicitly, are objectivist. I will discuss substantive list criteria many places along the way, and mainly in chapter 3, but here is summary account: A substantive list criterion is a criterion that employs an unchanging list of facts (which could themselves be subjective facts, but not subjective facts about actual agents, since these would be subject to change), which is referred to by the standards of correctness for the claims on the theory that employs it. For theories employing such a criterion, at least some facts are grounded on the elements of the list, otherwise it would be superfluous. And whenever a fact is grounded on such an element, that is really the only thing that the fact is grounded on. So on any theory employing a substantive list criterion at least one fact will not be grounded on subjective facts of actual agents, and hence would not be subjectivism.

Then there are the naturalist realist theories, some of which are reductive and some of which are not. The most famous form of non-reductive naturalist realism is what is often called Cornell-realism (after Cornell University) a view mainly developed and defended by Richard Boyd, Nicholas Sturgeon and David Brink (see for example Sturgeon, 1988; Boyd, 1988 and Brink, 1989). This type of view maintains that moral facts are a distinct type facts, just like non-naturalist objectivists do, but unlike the non-naturalist, the non-reductive naturalist thinks that moral facts are grounded by natural facts.¹⁶ Non-reductive naturalist realism is objectivist because not all the natural facts that ground moral facts are subjective facts.¹⁷ Another distinct form of non-reductive naturalist realism is Michael Smith’s anti-humean theory of normative reasons, (Smith, 1994: ch. 5). This theory is a bit like the sensibility theories, in that it builds much of its

¹⁶ Alternatively, that they *supervene* on natural facts or are *constituted* by natural facts.

¹⁷ Of course, a non-reductive naturalist realist *could* claim that the only natural facts that ground moral facts are subjective facts about actual agents, in which case the theory would be subjectivism, but I know of none who do.

framework around subjective facts, the main difference being that it aspires to naturalism. The key formulation of this theory is:

We have normative reason to ϕ in circumstances C just in case we would want to ϕ in circumstances C if we were fully rational. (Smith, 1994: 180–181)

Just like with the sensibility-theorists, the theory looks to ground normative¹⁸ facts in subjective facts, but includes a criterion that if fully unpacked will reveal itself to be a substantive list criterion. And as mentioned, all theories depending on such a criterion will be objectivist.

Finally, we have the reductive naturalist realists, who much like the non-reductive naturalists believe that moral facts are grounded by natural facts. In their case however, the moral facts are not only grounded by, but are also reducible to natural facts, meaning that we can provide a complete description of all moral matters without using any moral terms at all. Frank Jackson's moral functionalism is an example of a naturalistic reductive realism, which via a clever tool called "network analysis"¹⁹ aims to provide an analytic reduction of moral terms to natural terms (Jackson, 1998). Extremely roughly put, in a network analysis you make a long conjunction of all the relations that a moral fact stand in to other facts, moral and otherwise, and then replace all the moral terms in that list with variables (x , y , z , etc.). The long sentence that you get out of that (which is a completely natural fact, since all the moral terms are replaced) *is* the moral fact. Such a theory is objectivist, since what grounds moral facts on the theory just is all the relations that make up that long sentence, and those relations are objective facts. Now, not all reductive naturalist realisms are objectivist, as we will see.

In addition, we have some theories that don't easily fit into the labels of naturalism and non-naturalism, perhaps in large part because their proponents are hesitant to use the jargon common in metaethics. A lot of the theories of the Kantian school fit into this category. Any Kantian theory is likely to be objectivist, but the theory defended by Christine Korsgaard is an interesting example. Her theory, which is often labeled "Kantian constructivism" (Bagnoli, 2015), is another objectivist theory that employs many of the tools of the subjectivist. Simplifying the matter to an extreme degree, Korsgaard's theory tries to account for what one ought to do in terms of what an agent must want to do insofar as that agent wants to be a rational agent that is a member of the community of rational agents (i.e. the Kantian notion of the Kingdom of Ends), (Korsgaard, 1996: chs. 3 and 4). But this, of course, assumes that we do want

¹⁸ That we in this case talk about normative rather than moral facts makes no difference in the current context.

¹⁹ Which is a type of Ramseyfication. See for instance Smith (1994: 44–47).

to be rational agents that are members of the community of rational agents, and so, to be successful, the theory needs this to be a requirement, and the requirement to want to be a rational agent that is a member of a community of rational agents is another example of a substantive list criterion, making Korsgaard's theory objectivist.

There are many more objectivist metaethical and metanormative theories, but I hope this has given you an idea of the broad range of possible objectivist theories. The conclusion I argue for in this thesis is only that any metaethical or metanormative theory must be objectivist, so any of the theories presented in this section could potentially be correct given that conclusion. Now, it is time to move on to subjectivism.

1.5 Why Subjectivism?

As we discussed above, it might seem strange to think that morality should be a lot like the domain of favorite colors, intuitively speaking. Using Enoch's three tests, we (hopefully) elicited a strong feeling that morality *should* be objective. But of course, those tests were carefully rigged to elicit those intuitions, and if we try to ask some different questions, we might get some intuitions that go the other way. For instance, there is a common picture on which certain sciences are *hard* and others are *soft*. The natural sciences, like physics, are hard on this picture. Humanistic sciences like art history or esthetics are on the softer end of the spectrum. Much of what is meant by the hardness of the natural sciences or the softness of esthetics can be cashed out in terms of the objective/subjective-distinction, hard corresponding to more objective while soft corresponds to more subjective. If we try to place the discipline of ethics onto this continuum, it will probably be placed towards the softer end. Morality, after all, feels quite a lot unlike the natural sciences.

So, is morality objective or subjective? I predict that a layman's answer to the question of whether morality is more like favorite colors or natural sciences would often be something like, "it's a bit of both." So a radical subjectivism that makes the criteria for moral rightness too much like the criteria for being a favorite color would be rejected by most people. This is why most subjectivist theories look like they contain some elements of objectivism. It would be embarrassing for any metaethical theory if the right answer to a moral question like "is abortion wrong" will vary depending on whether you ask me, or the person sitting across from me right

now. One way a subjectivist theory could avoid this kind of seeming arbitrariness is to include reference to more than one individual in the criteria for rightness. Another way is to include some counterfactual criterion. I will discuss all of these strategies, and what kind of theories they result in, below. For now, the important thing to note is that these theories can still be called subjectivist in the most general sense if the criteria for moral rightness etc. that they produce are similar, in an important and fundamental sense, to the criteria for being a favorite color. If, for instance, the criterion for rightness depended on what the people in a society collectively feel about the case, the answer to the question “why is murder wrong?” might be something like “nearly all of these people in this society dislike murder.” And that is a lot like “because I just *like* red.” The kind of criteria that a subjectivist can add to her theory to make it more likely to be intuitively appealing are what I call *neutral function criteria*. The full explanation of these will have to wait until chapter 4 (see also section 2.3.3), but the important thing to note about them is that they are criteria that allow the subjectivist to narrow down the range of intuitively weird consequences of their theory, without making the theory objectivist. They are therefore similar to substantive list criteria in terms of their purpose, but with drastically different consequences for the theories that employ them.

In addition to the intuitive concerns already mentioned, there are quite a few philosophical arguments that support the push for subjectivism. One classic argument is one of the main motivations for *relativist* metaethical theories, like Gilbert Harman’s framework relativism (Harman & Thomson, 1996). This is an argument one can construct from the empirical fact that people have differing opinions concerning moral claims (both across cultures and individuals), and that some differences of moral opinion cannot be resolved by any amount of rational argument. If successful, this argument would make us think that matters of morality are somewhat like matters of favorite colors, where we would take it as a matter of course that we might never be inclined to change our favorite color, regardless of the amount and quality of rational argument, for what kind of argument could possibly persuade you to change your favorite color?²⁰ Although there are unresolved debates in the natural sciences, the disagreements they concern are normally thought to be resolvable in principle, and, importantly, what will

²⁰ There are situations where it would be rational to change one’s favorite color in response to some argument, say an argument like, “unless you change your favorite color to red I will shoot you in the knee.” But this does not seem like the relevant kind of argument, even if you assume that it is possible to voluntarily change one’s favorite color.

resolve the disagreement will be paradigmatically objective things such as further empirical evidence. Is there any such evidence is available to the parties disagreeing about a moral issue? There are no obvious candidates. This point about irresolvable disagreement be construed both as an argument for moral relativism and for metaethical subjectivism, and I take most interesting variants of moral relativism to *be* species of metaethical subjectivism.

Another ubiquitous argument for subjectivism has to do with moral motivation. It is often taken to be an essential requirement of any good metaethical theory that it gives a satisfactory account of how moral judgments can motivate us to act. Michael Smith calls this the “practicality requirement,” and formulates it as follows:

If someone judges that it is right that she ϕ s then, *ceteris paribus*, she is motivated to ϕ . (Smith, 1994: 12)

Although a similar type of concern is voiced by many authors, Smith has one remarkably clear way of formulating the problem, which is the eponymous main subject of his book *The Moral Problem*. For Smith, the problem presents itself once we add two more premises: The first is that to make a moral judgment of the form “it is right for me to ϕ ” just is to form a belief about an objective moral fact of the form “it is right for me to ϕ .” The second is what he calls the Humean theory of motivation, which tells us that agents are motivated just in case they have an appropriate desire and a means-ends belief, but that there is no necessary connection between having a certain belief and having a certain desire. Together, these three premises form an inconsistent set. The practicality requirement together with the definition of what it is to form a moral judgment tells us that whenever I form a belief like “it is right for me to give to charity,” I am, *ceteris paribus*, motivated to give to charity. But this requires there to be a necessary connection between the formation of a certain belief and having a certain desire, which the Humean theory of motivation tells us is not the case. So, to solve the problem, one of the premises has to go. Smith’s solution is to abandon the Humean theory of motivation in favor of his anti-Humean theory of normative reasons described in section 1.4. His thought is that the subjectivist elements of this account of normative reasons can plausibly account for the strong link between having a belief and being motivated to act on it. So even though Smith’s theory is ultimately objectivist, it is the subjectivist resources he makes use of that in the end solves the problem.

Indeed, it is not hard to see why subjectivism could provide resources for solving the problem of moral motivation. After all, subjectivists claim that moral facts are grounded in

subjective facts, a good number of which are what philosophers like to call pro- and con-attitudes, which are what we might call motivational attitudes. On some understandings of motivation, being motivated to ϕ just is having a pro-attitude towards ϕ -ing. The tight connection between moral facts and pro- and con-attitudes, then, helps explain the tight connection between making moral judgments and being motivated. The main reason why Bernard Williams is a subjectivist (1981: ch. 8), is that he thinks that any claim about what an agent has reason to do that is not grounded in the subjective facts that are true about that agent just could not possibly motivate the agent. Since he thinks that all reasons must at least be potential motivating reasons, this means that all reasons must be grounded by subjective facts about the agent (1981: ch. 8). Harman (2000) and Jesse Prinz (2007) make related arguments for their respective versions of relativism.

The third argument in favor of subjectivism is an argument from epistemology, which is can mainly be considered an objection to the Platonist versions of objectivism, where there is supposed to be a special problem about accounting for how we, in metaphorical terms, can have access to this Platonist realm of moral facts. More precisely, the problem can be put in terms of how we can have justified beliefs about facts that are outside the realm of the natural. All of our other senses work through causal interaction with other parts of the natural realm, but Platonic moral facts seem to be completely closed off from this (Platonic facts are usually thought to be acausal). Of course, the name of intuitionism hints at the preferred solution for a large group of the Platonist theories: the special faculty of moral intuition, which is a faculty that can grasp truths about the Platonic realm of moral facts. But again, due to the considerations about the acausal nature of the thing that is to be grasped, it seems mysterious how this faculty is supposed to work. There is a close analogy to the epistemological problem for moral Platonism that applies to the theory of mathematical Platonism (indeed it looks a lot like the very same problem). In mathematical Platonism, mathematical objects like numbers are acausal Platonic objects. Hartry Field presents the challenge as follows²¹: Mathematicians are really good at getting the answers to mathematical questions right. There seems to be a reliable correspondence between our beliefs about mathematical facts, and actual mathematical facts. But if mathematical facts are acausal Platonic entities, there seems to be no way of *explaining* this reliable correspondence. After all, there can be no causal interaction between the mechanism for forming the beliefs about the

²¹ This presentation of Field's argument is based on David Enoch's presentation of it in his (2011: 158–159).

mathematical facts and the facts themselves. The exact same problem applies to the moral and normative case. This seems to be quite a significant problem, and many take it to be a decisive reason to reject Platonism about both mathematical and moral (or normative) facts (Enoch 2011: 159; Mackie, 1977: 38–39). Even some convinced Platonists like Scanlon, think that the epistemological challenge is serious enough to weaken his confidence in his theory (2014: 122). Sharon Street’s “A Darwinian Dilemma for Realist Theories of Value” (2006) can be seen as a special version of this epistemological challenge that applies with force to moral Platonism (Enoch, 2011: 163–165). I will discuss Street’s challenge in more detail in chapter 4 as it will be particularly relevant as an objection to one of the argumentative strategies in this thesis.

The epistemological challenge is, as noted, mainly an argument against Platonistic versions of objectivism. Naturalist versions will not run into the same problem of explanation, because there the moral facts are natural facts and hence not acausal. With regard to sensibility theories, the case is a little less clear, but my gut feeling is that the epistemological problem will be equally hard for them to answer, since the substantive list criterion on a non-naturalist theory, will require the existence of some objective non-natural fact. And this, I think, is really just Platonism in another guise.

The fourth type of challenge to objectivism are those from ontology and metaphysics. These, too, challenges directed against the non-natural and mainly the Platonist versions of objectivism, and they mostly concern the extra bits that have to be added to our ontology in order to make room for moral facts. Speaking in terms of parsimony, a theory that postulates an entire realm of a completely distinct type of facts could seem extravagant indeed. An analogous case is that of monism versus dualism in the philosophy of mind. To postulate that there are two entirely distinct types of fact, physical facts and mental facts, just seems unparsimonious, at least if there is a plausible alternative. Still, there have been numerous often quite vigorous defenses of dualism in recent years (e.g. Chalmers, 1996), and it seems like parsimony, by its very nature, is a constraint that can be quite easily overridden by other concerns. Parsimony is a highly defeasible criterion, one that comes into play most forcefully when we have two theories that are equally good in other regards. So, if we have a good reason to suppose that Platonic entities need to exist, this could be overridden.

There is a type of argument that has most famously been run in favor of the existence of Platonic mathematical facts, that can be run for the existence of normative facts (although the

arguments will considerably different content in the two cases). This the “argument from indispensability.” Without going into any sort of detail, an indispensability argument argues that some contested thing is indispensable for some other project that we already believe in or require to be possible. In the case of mathematics, the Quine-Putnam indispensability argument argues as follows: Anything that is indispensable to our best scientific theories deserves a place in our ontology. Mathematical facts are so indispensable. Therefore, mathematical facts deserve a place in our ontology (Colyvan, 2015). David Enoch has a similarly modeled indispensability argument for irreducibly normative facts. He argues: Anything that is indispensable to a project that is itself intrinsically indispensable, deserves a place in our ontology. The deliberative project is intrinsically indispensable, and irreducibly normative facts are indispensable for the deliberative project. Therefore irreducibly normative facts exist (Enoch, 2011: ch. 3). If any of these arguments are successful, it would seem like they could easily overcome the parsimony requirement.

There are further metaphysical worries with regards to moral facts in general, and Platonism in particular. Here, the lines will start to blur between what are genuinely ontological concerns, and what are more general metaphysical issues. In any case, the most famous of these concerns is J. L. Mackie’s *Queerness-argument*. Here is Mackie: “If there were objective values, they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe,” (1977: 38).²² The first part is a metaphysical worry, expressed by the quote above. As it stands, it is a bit hard to know what to make of this challenge, other than the fact that Mackie thinks objective values are weird. To be fair to Mackie, his text includes many things that hint at possible interpretations of his queerness-claim that could correspond to any of the other objections to objectivism that we have discussed so far, and the ones we have yet left to discuss. Yet the queerness objection just doesn’t seem to do anything interesting all by itself. As Platts put it,

[t]he queerest thing about this as it stands is the claim that it is an *argument*. ... The world *is* a queer place. I find neutrinos, aardvarks, infinite sequences of objects, and (most pertinently) impressionist paintings peculiar kinds of entities; but I do not expect nuclear physics, zoology, formal semantics or art history to pay much regard to that. (Platts, 1980: 72, quoted in Enoch, 2011: 135)

²² Mackie originally presented his queerness challenge in two parts, one of them being the epistemological challenge discussed above. But it is the metaphysical worry that is most commonly associated with the argument from queerness.

One way to understand the argument and give it a bit more force is to think of it as a trumpet blown against non-naturalism in general. These days it is very common aspiration of philosophers to be naturalist. If you are not a naturalist, then you might be considered backward, or disrespectful to the sciences. Or, at least, I take it that this is what many philosophers think. It is somewhat more tricky to spell out what exactly naturalism is, and why we should be so eager to label ourselves as such. Naturalism, in its clearest formulations, is commonly split into two types (e.g. Leiter, 2002: 3; Miller, 2013: 3; Papineau, 2009):

Methodological naturalism: The view that philosophy and other inquiries aimed at truth should proceed as an extension of the natural sciences, emulating their methods and taking their discoveries as a starting point.

And

Ontological naturalism: The view that the kind of entities that the natural sciences postulate or are likely postulate in the future are the only things that fundamentally exist.

When I labeled theories non-naturalist in section 1.4, the naturalist part refers to ontological naturalism, so a non-naturalist theory is one that rejects ontological naturalism, but not necessarily methodological naturalism. Now, the queerness argument could be a way of saying that any type of objective moral facts are a kind of thing that the natural sciences are not likely to postulate. This seems like a pretty reasonable assumption. And on that assumption, anything that postulates objective moral facts will be in conflict with ontological naturalism. But we know from before that there are naturalist theories that also postulate objective moral facts. These theories, then, need to argue that objective moral facts actually are a kind of thing that science is likely to postulate. And this is exactly what they do.

Of the two versions of naturalism, methodological naturalism is the one that seems most respectable. Not being a methodological naturalist does indeed seem a bit backward and disrespectful of the tremendous success of science. But it is ontological naturalism, not methodological naturalism, that is violated by non-naturalist theories, and it is ontological naturalism that is needed for the queerness-argument. This could be bad news for the queerness-argument, since to lend ontological naturalism the same credence as methodological naturalism seems a bit fishy. After all, as David Enoch points out, there is no “and that’s all”-clause to

science as a whole, so ontological naturalism is in no way implied by science itself (apart from the kind of parsimony consideration we discussed above) (2011: 135). So, I find the argument from ontological naturalism quite unpersuasive.

There is, however, a much more sophisticated and deep metaphysical challenge to Platonist objectivism, which originally comes from Simon Blackburn: the argument from the supervenience of the normative on the natural (Enoch, 2011: 142; Miller, 2013: 28–30; Blackburn, 1973).²³ The supervenience argument is quite complicated, and I will confess beforehand that I am not really sure I understand it (at least I cannot seem to find it as compelling as other people seem to find it). Different philosophers interpret it differently, and sometimes it seems that they disagree completely about what makes the argument work. However, I will *try* to give it a fair presentation of at least one interpretation of the argument. This is the “Ban on Mixed Worlds” version of the supervenience argument (Miller, 2013: 48–52). I hope I may be forgiven if I fail to do the argument justice.

Normative facts supervene on natural facts, and all normative properties supervene on natural properties—that is, if some *x* and some *y* have all the same natural properties, then they must have all the same normative properties as well. This is a conceptual or logical truth.²⁴ But natural facts do not seem to *necessitate* moral facts (we are assuming that naturalism is false here), because “[i]t does not seem a matter of conceptual or logical necessity that any given total natural state of a thing gives it some particular moral property,” (Blackburn, 1984: 184, quoted in Miller, 2013: 50). The argument we are going to present is trying to capitalize on the different consequences that necessitation and supervenience has, and the explanatory burden this creates for the Platonist.

If it were the case that natural properties both necessitated *and* supervened on moral facts, the following would be true: For a given moral property *M*, there is a natural property *N*, where *x* is *M* just in case *x* is *N*. On this view, the following is *not* possible: To have a world *W*,

²³ There is another powerful supervenience argument against Platonist theories due to Frank Jackson, but as this argument is an argument against one type of objectivism (Platonism) for another type of objectivism (reductive naturalist objectivism), I will not discuss it here. For the argument, see Jackson (1998). For a response on behalf of the platonist, see Enoch (2011: 137–140).

²⁴ If you doubt this, try to picture two completely identical sets of natural circumstances, say a man torturing a dog, and an identical man torturing an identical dog in the exact same way, with all the circumstances and history around the event also being identical. Could you imagine that the one man torturing the dog was doing something right while the other was doing something wrong? If you think you can, check whether you also imagined in some difference in a natural fact, like some important thing that the torture would accomplish.

where there is a *b* that is both N and M, and to have a different world *W'*, where there is an *a* that is both N and not M. In other words, it is not possible to have two distinct worlds with identical natural properties but different moral properties. This seems to be just how it ought to be.

Now let us look at what is the case if, as we argued, supervenience is true but necessitation is *not*. Now, it is no longer the case that *x* has moral property M just in case *x* has natural property N. On these assumptions it is completely (logically) possible for there to be a world *W* where *b* is N and *b* is M as well as a distinct world *W'*, where *a* is N but not M. But what is *not* possible, is for there to be a single world, in which there is a *b* that is N and M as well as an *a* that is N but not M. We can say that supervenience creates a *ban on mixed worlds*, i.e. worlds where two worldmates have the same natural properties but different moral properties.

What is the problem with this? While it might not be immediately obvious, there is a fact that begs to be accounted for here. Remember that we argued that supervenience but not necessitation is true, so on these assumptions it will be the case that no worldmates can share natural properties without sharing moral properties, but counterparts across possible worlds can be identical in all natural respects without sharing moral properties. And this is a fact that cries out for explanation. Why must identical worldmates share moral properties, when identical counterparts in distinct worlds do not? The conceptual truth of supervenience creates, or is, something that we need to explain.

The next step of the argument is to argue that Platonists will have an especially hard time coming up with such an explanation. For on their theory moral facts and natural facts are completely distinct existences, normative facts being acausal entities. And what the ban on mixed worlds needs us to explain is that, within a given world, there is a perfect lawlike ordering of normative facts and natural facts. How can there be such a perfect lawlike ordering between natural and normative facts when normative facts are acausal entities that don't interact with natural facts? No explanation seems forthcoming.

For subjettivists, and other naturalists, the answer is obvious, and the problem seems not even to arise. After all, if moral properties just are, or are reducible to, natural properties, we could say that natural properties necessitate moral properties after all. Or, if we still want to avoid talk of necessitation, we can have a completely straightforward explanation for supervenience and the ban on mixed worlds. For on naturalism there is a perfectly ordinary casual relationship that underlies and explains the lawlike ordering of natural and moral facts,

because on naturalism, moral facts just are natural facts. And relations between natural facts and other natural facts are not particularly problematic to explain, at least not on this abstract level.

Now, it seems to me that this problem has an obvious and completely satisfactory solution.²⁵ Natural facts do not, indeed, necessitate moral facts. However, moral facts are *themselves* necessary. This can be readily realised if we take moral facts to be a priori (Kripke notwithstanding).²⁶ And if this is the case, the problem of the ban on mixed worlds disappears, because there simply is no possibility of two distinct worlds with identical natural properties but differing moral properties. Since moral facts are necessary facts, they are true on all possible worlds. It also explains the lawlike ordering of moral and natural facts, if we look on necessary moral facts as analogous to other necessary truths, most of which are *lawlike*, like logical and mathematical truths. These do not have any causal connections to other types of facts, but they have perfectly ordered and lawlike relation to natural facts, say. The nature of the relation is one of a law to those bound by it. There is no causal interaction going on between the law (in the literal, not the metaphorical, sense) itself and its subjects, but there is a perfectly ordered relation. And moral facts are just like this. Just think of Kant's categorical imperative, or the principle of utility. This, to my mind, solves the problem, but if my answer fails, the supervenience argument will be one more argument that could favor subjectivism.

1.6 Summary and Preview

In this chapter I set out to define subjectivist theories of ethics and normativity versus objectivist theories about these domains. Subjectivist theories are those that whose standards of correctness always refer to subjective facts about actual agents, or whose claims always are grounded by subjective facts about actual agents. Objectivist theories are those that fail to be subjective on this definition.

I provided three tests designed to elicit objectivist intuitions before I went ahead to give an overview of the metaethical landscape of objectivist theories. Importantly, I mentioned how

²⁵ I have a small suspicion that I have simply misunderstood something rather than the solution being good, because the solution just seems a bit too obvious not to have been discovered earlier. For the life of me though, I cannot figure out what is wrong with it.

²⁶ I also happen to think that this is the only way to make sense out of the faculty we have for grasping truths about normativity. It is a faculty that allow us to grasp a priori truths, the same type of faculty (perhaps even the same faculty) that allow us to grasp truths about mathematics and logic.

some theories making use of what we might call subjectivist resources ended up being objectivist Platonist theories. I then proceeded to make a case for why you might want to be a subjectivist.

Summarizing this case, there are, as with objectivism, some intuitive grounds for being a subjectivist. These grounds do seem at least a little bit weaker than the intuitive grounds for being an objectivist. To make up for this, there are a host of arguments against objectivism (Disagreement, Moral Motivation, Epistemology, Ontology and Metaphysics). And while only the argument from persistent disagreement is an argument against all types of objectivism, they will together build a fairly strong case for subjectivism. This is especially true since the two main thrusts of the challenges complement each other. The challenge to accommodate moral motivation is one that pushes theories to embrace subjectivist resources, which is what sensibility theory, Korsgaard's Kantian constructivism and Smith's anti-Humean theory of normative reasons do. But these theories, since they employ substantive list criteria, are best understood as Platonism in disguise. That makes them vulnerable to the second thrust of the objections: the challenges that target Platonism. These are the epistemological challenges and the ontological and metaphysical challenges. All in all I believe we have a solid case for thinking that subjectivism is a strategy we ought to explore. In the following chapter, I will start exploring a particular argumentative strategy against subjectivism. I will then start to present particular subjectivist theories, which I will attempt to attack with using the argumentative strategy. In response to each attack I will present a new and increasingly sophisticated versions of subjectivism, designed to cope with the attack. In the end however, I argue that the only resource the subjectivist can successfully employ against the arguments will be a substantive list criterion, i.e. one that will make the theory objectivist.

2. Moorean Arguments Against Subjectivism

In this chapter I will present what I think are some of the most important arguments one can wield against metaethical subjectivism. It is a family of arguments that I call Moorean, since they are in the spirit of G. E. Moore's famous "two hands" argument in "Proof of an External World"

(Moore, 2008). I will then start to develop some naïve versions of subjectivism and show how Moorean arguments apply to each of them. The way that the theories are vulnerable to the arguments will show us how they need to be improved.

2.1 Basic Moorean Arguments

Moore's original argument, which I will call HANDS, goes like this (Moore, 2008: 26):

HANDS	P1: If the external world does not exist, I do not have hands.
	P2: I have hands.
	C: The external world exists. (By <i>modus tollens</i> .)

The argument is strikingly simple and may even seem a little silly. Despite this, it is philosophically deeply interesting, and it seems to have some considerable force.²⁷ After all, which premise are you supposed to question? P1 is a straightforward statement of an implication that the nonexistence of the external world has—there is nothing to doubt here. P2 is commonsensical and obvious, and that is the point. There is nothing extremely special about *hands* specifically: with a corresponding adjustment to P1, I could have chosen any physical object whose existence I am as confident in as the existence of my hands as material for P2.²⁸ Still, P2 is the crucial premise of this argument, and it is the one that ought to be questioned. To see why, it will be useful to look at the wider dialectic in which the argument is supposed to function.

The conclusion of HANDS is commonsensical. Almost no one doubts that the external world exists—no one, that is, apart from those who are skeptics about the existence of the external world. The argument would therefore not be relevant unless it is meant as a rebuttal to arguments *against* the existence of the external world, and that is of course precisely what it is. So, the dialectic goes something like this: (1) An argument is presented that has the conclusion

²⁷ Many prominent philosophers endorse it, and, as I will talk more about in chapter 4, there has been a recent resurgence of interest in both Moorean arguments in general and HANDS in particular. In his (2009: 1) Tristram McPherson mentions David Armstrong, Kit Fine, Thomas Kelly, David Lewis, William Lycan and James Pryor as examples of some prominent supporters of the argument.

²⁸ Though, once we look a bit closer and start to really question the argument, as we will in chapter 4, we shall see that there actually is something about your hands that make them especially suited for the argument, namely the way that you are so intimately acquainted with them through how you use them to interact with the world.

that the external world does not exist. (2) HANDS is presented as a rebuttal. By starting with the conclusion that the external world does not exist, we can set the Moorean argument up like a *reductio*:

HANDS	1.	The external world does not exist. (From external world skepticism)
(REDUCTIO)	2.	If the external world does not exist, I do not have hands. (Linking premise)
	3.	I have hands. (Moorean premise)
<hr/>		
	4.	The external world exists. (By <i>reductio</i> from 1., 2. and 3.)

The *reductio* version of the argument does not do or require anything more than the original *modus tollens* version—this is an important point as I will use the *reductio* version later—but it does make what is going on a bit clearer. To really make it clear what is going on, I have labeled each of the three premises according to their function in the argument. First we have the *reductio assumption*, which comes from the content of the theory that the argument targets. This is the most straightforward of the three premises—nothing surprising is going on here. Second, we have the *linking premise*, which is an assisting premise that provides the logical link between the *reductio assumption* and the following *Moorean premise* required to generate the contradiction. The linking premise aspires to be a fact that by itself is neutral to the truth or falsity of both the *reductio assumption* and the *Moorean premise*, meaning that the linking premise should come out as true regardless of whether it is the theory that is being attacked that is true or the *Moorean premise* that is true. Last, but certainly not least, is the *Moorean premise*. This is supposed to be a fact that, through the linking premise, contradicts the *reductio assumption* and so the theory that is being attacked. The *Moorean premise* is the driving force behind any *Moorean argument*, and it is the one part of the argument that is really open to serious questioning. But, precisely because of this, any successful *Moorean premise* is supposed to reflect what we will call a *Moorean fact*. And a prime characteristic of a *Moorean fact* is that it is a fact that it is really hard to question.²⁹ The fact that I have hands is one such fact.

It might look like the *Moorean argument* is question-begging. After all, what the claim that the external world does not exist really amounts to is the claim that no such things as hands and tables and the like exist, so, when we include the existence of hands as a premise, it looks

²⁹ I will give closer scrutiny to what exactly it takes to be a *Moorean fact* in section 4.2.2.

like we are simply denying what the argument is supposed to disprove at the outset. In a sense, this is what the Moorean argument *is* doing: it is asserting that the external world does exist by pointing to facts we are extremely confident in that entail just that. Another way to put it is that the Moorean argument is showing us—beating us over the head with—the beliefs that we have to give up if we embrace the conclusion of the skeptical argument. I actually have to give up the belief that I have hands! And before I give up that belief it might be rational to see if there are any of the premises that led to the conclusion that the external world does not exist that are weaker, and so less costly to give up, than the belief that I have hands. One common way of evaluating the strength of a Moorean argument is therefore to parade it next to an argument in favor of the theory that you are attacking, and compare the Moorean premise to each of the needed premises in this argument, in order to see whether any of them are weaker than the Moorean premise itself (Huemer, 2005: 115–117).

The most obvious use we can put Moorean arguments to in metaethics is against arguments for nihilism or moral error-theory. Nihilism is the view that nothing is good or bad, or right or wrong, and moral error-theory is the view that moral statements are systematically false. These views go together naturally, and have overlapping consequences, but since error-theory is the only view that is a major player in the metaethical field today, I will speak only of the latter. Here is a Moorean argument against error-theory:

- | | |
|----------|---|
| NEIGHBOR | P1: If moral statements are systematically false, it is false that it is wrong for me to go and kill my neighbor right now. |
| | P2: It is true that it is wrong for me to go and kill my neighbor right now. |
| | C: Moral statements are not systematically false. (By <i>modus tollens</i>) |

P2 seems like a really strong premise; to give it up would be very costly. It looks like any error-theorist will, in light of the possibility of this kind of argument, be fighting an uphill battle. We could do the comparison test mentioned and compare NEIGHBOR with a common argument for error-theory (you might recognize this from our discussion of the metaphysical challenge against Platonism):

- | | |
|-------|---|
| QUEER | P1: If moral facts exist, they would be entities of a very strange sort, completely unlike anything else that we know of, i.e. they would be queer. |
| | P2: Queer entities do not exist. |

Are any of the premises in QUEER weaker than the Moorean premise of NEIGHBOUR? If so, it looks like we have reason to reject that premise in favor of the Moorean one.³⁰

Employing Moorean arguments against error-theory is interesting, and I believe that my discussion in section 4.2 and its subsections could shed some light on the ultimate fate of such attempts.³¹ But error-theory is not our main opponent here. The main reason for presenting this argument here was to show how the classical Moorean style of argument can be applied to the case of metaethics. In this case, the situation is very similar to that which we had with HANDS. The Moorean argument is presented as a response to what might be called a skeptical theory about some type of thing—external objects in the first case and moral facts in the other. In both cases, it looks like the strength of the Moorean argument rests on how strong its Moorean premise is.

Below I will show how we can formulate Moorean arguments against subjectivist theories. These arguments will have the exact same basic form, but they will perhaps feel slightly different. This is mainly because the theories we are employing them against are not skeptical theories, but rather what I will call *revisionary* theories. Before I present the subjectivist theories and attack them, I will say a bit more about how the Moorean arguments are going to work in this context and discuss what it means for a theory to be revisionary

2.2 Revisionary Theses and the Limits of Moorean Arguments

The targets of the arguments in the previous section were theories that entailed skepticism about the existence of some type of thing. We might call these kinds of theories *skeptical*³² theories. One might think that Moorean arguments are mainly suited to attack such skeptical theories.

³⁰ This is not really intended to be a comprehensive discussion of Mackie's queerness-argument (my task here is not to battle error-theory), and so I am perhaps not presenting the error-theoretical case with as much force as it could be presented. Still, many philosophers, I take it, are convinced by exactly this kind of consideration. Jonas Olson mentions A. C. Ewing as an early proponent (Olson, 2014: sect. 7.1). See also (Fine, 2001) and especially Huemer (2005: 115–117).

³¹ Okay, let me say just a little bit more: I do not necessarily think that Moorean argument are a decisive move against moral error-theory. But I do think they represent a problem that error-theorists *have* to answer. There are a number of things error theorists can do to do just this, and the most important of them I discuss in section 4.2.3.

³² "Skeptical" is here used to denote theories that make claims about the nonexistence of things, not just the lack of knowledge about the existence of things, i.e. in an ontological rather than an epistemological sense.

However, skeptical theories are a subclass of a broader kind of theory that we might call *revisionary*. Revisionary theories are theories that make claims that require us to revise our common sense views about the things the theory is about. Revisionary theories stand in contrast to *descriptive* theories, which are theories that try to provide descriptions that match our common sense views. Revisionary theories come in degrees. Most theories are revisionary to some degree—*mildly* revisionary—perhaps requiring us to adjust or abandon some of our least important common sense beliefs. Other theories are *radically* revisionary, perhaps requiring us to abandon most of our strongly held common sense beliefs. The factors that take a theory from being mildly revisionary to radically revisionary vary along a few axes: the number of beliefs that have to be altered, and the strength and centrality of those beliefs.

The *strength* of a belief amounts to how confident we are in the belief, and the *centrality* of a belief is how crucial the belief is to the domain of discourse that is being revised. Say we are discussing a revisionary theory about kitchen utensils, and one of its consequences is that we have to abandon our belief that spoons are a suitable tool for eating soup. This is certainly a very strong belief, and abandoning it would be somewhat costly. Nevertheless, it is conceivable that due to the theories' other virtues we might come to accept it despite having to abandon this strong belief, especially if this was the only belief that we had to abandon. The more strong beliefs that we would have to abandon, the more radically revisionary the theory is. Another theory about kitchen utensils might require us to abandon our belief that kitchen utensils are tools that are, among other things, meant to be used in the preparation of food. This belief is central to the discourse about kitchen utensils in such a strong way that if we were to abandon it, someone could rightly question whether or not we still were talking about kitchen utensils and not something else entirely. Such a theory would be radically revisionary indeed. Both strength and centrality come in degrees. A central belief is almost always also strong. Abandoning a few merely strong beliefs is usually less radical than abandoning a single central belief.

What I previously called skeptical theories will fall on the extremely radical end of the spectrum of revisionary theories. Both external world skepticism and moral-error theory call for the abandonment of beliefs that are extremely central to the domains they are theories about—the existence of external things, the truth of moral statements. Moorean arguments are responses to arguments for revisionary theories. The more strongly, the more radically, revisionary the theory is, it seems, the more likely it is that Moorean arguments can be wielded against it. It seems clear

that the scope of theories that Moorean arguments might be wielded against is very broad—perhaps too broad to make it likely that Moorean arguments can succeed in every case. After all, it does not seem plausible that there is a blanket restriction to the degree that revisionary theories can be successful that such broad applicability of Moorean arguments might suggest.³³ We do think that some radically revisionary theories are successful, despite the possibility of Moorean arguments. Take what we might call error-theory of witches. This theory states that no witches exist and no spells are efficacious, and, back in the day, the theory would count as revisionary because the commonsense view was that witches existed and that spells worked. The theory would even be radically revisionary, because the beliefs that the error-theory asks us to abandon are *the* central beliefs in the witches-and-spells discourse. We might thus imagine a witch-believer constructing the following Moorean argument:

WITCH	P1: If no witches exist, then my grandmother was not a witch.
	P2: My grandmother was a witch.
	C: Witches exist. (By <i>modus tollens</i> .)

This seems like an argument that no one in their right mind would take seriously. But it has precisely the same structure as both HANDS and NEIGHBOR, which I discussed above. Now, we might think that WITCH is no good as an argument simply because unlike in HANDS and NEIGHBOR, the Moorean premise of WITCH simply is not true. It is not a Moorean fact, because it is not a fact at all. But this fact is only apparent to us because we live in a time when witch-belief is almost eradicated. What if there is a future where belief in external things is almost eradicated and from whose perspective the Moorean argument against skepticism about the external world would seem just as ridiculous as the argument for the belief in witches. “Imagine using ‘I have hands’ as a premise!” they might say, “obviously, there are no such things as hands.” Although, to us, this seems far-fetched, it does look like it is possible.

To see that this type of aberrant Moorean argument is not restricted to such antiquated domains as that of witches and magic spells, we can consider the following Moorean argument, from Tristram McPherson’s paper “Moorean Arguments and Moral Revisionism,” (2009: 14–15). The argument goes like this:

³³ Although, see Fine (2001), and the discussion in section 4.2.1 below.

BRICK	P1: If contemporary physics is true, then bricks are made up of mostly empty space.
	P2: Bricks are not made up of mostly empty space.
	C: Contemporary physics is false. (By <i>modus tollens</i> .)

Contemporary physics may be seen as a revisionary theory. Many of our common sense beliefs, such as the one that bricks are not made up of mostly empty space, must be abandoned if we are to accept it, but few people would take BRICK to be a serious challenge to the credibility of contemporary physics. If WITCH and BRICK are genuine examples of Moorean arguments, then clearly, Moorean arguments are not applicable to revisionary theories across the board. I think both WITCH and BRICK *are* examples of Moorean arguments, but that they are examples of *bad* Moorean arguments. The difference between good and bad Moorean arguments lies in the strength of their Moorean premises. Not all Moorean premises reflect genuine Moorean facts, and some Moorean facts are more *robust* than others. In section 4.2.2, I will give a more complete account of how to evaluate the force of Moorean arguments. The important thing to note for now is that there is some reason not to simply take Moorean arguments at face value—they do need to be individually evaluated so that we do not end up mistaking a bad Moorean argument for a good one. Because of this fact, I will postpone the verdict of whether the Moorean arguments I present against each form of subjectivism in this and the next chapter are to count as arguments against that theory on their own until the last section of chapter 4 (if you are really impatient for an answer, you have my permission to skip ahead.)

Perhaps you will now ask: “But if we are not sure whether or not the arguments will count, why do we even bother to see if we can formulate them?” The answer is that there is a reason to care about whether or not we can form at least *seemingly powerful* Moorean arguments against the subjectivist theories, a reason which is independent of whether or not the Moorean arguments will turn out to be *forceful* in their own right. I will take a Moorean argument to be *seemingly powerful* just in case it has the face value “oomph” of a good Moorean argument. Being seemingly powerful is logically independent to what I call being *forceful*, which is what it takes for an argument to on its own count in favor (epistemically speaking) of its conclusion.³⁴ So, what I need to show in the following sections is that we can formulate at least seemingly

³⁴ Alternatively, “to be an epistemic reason to believe in the conclusion.”

powerful Moorean arguments against the subjectivist theories. The full argument for why this is important comes in section 4.1 and its subsections (again, feel free to skip ahead), but the gist of it is this: the possibility of forming seemingly powerful Moorean arguments against a theory is enough (or so I argue) to identify that theory as radically revisionary. But subjectivism aspires to being only mildly or moderately revisionary, so an argument that shows it to be radically revisionary will itself be a blow to the theory. If we can argue further that morality and normativity are domains about which we should be hesitant to accept radically revisionary theories, we can see that we have some reason to reject subjectivist theories. Now, onto the theories and the arguments.

2.3 Moorean Arguments Against Metaethical Subjectivism

As defined in chapter 1, metaethical subjectivism is the view that the standards of correctness for all moral claims must refer to subjective facts of actual agents, or that all moral facts must be grounded by subjective facts of actual agents. There are numerous possible ways of cashing out this broad metaethical view into specific metaethical theories, but I believe that all theories that fall under metaethical subjectivism are equally vulnerable to a type of Moorean arguments. These arguments might feel and look slightly different to the classical Moorean arguments like HANDS and NEIGHBOR mainly because of the different type of theory that they target. The original target of classical Moorean arguments was skepticism of one sort or another, and these are radically revisionary theories that are completely explicitly so. Subjectivist theories, on the other hand, are often not even necessarily meant to *be* revisionary. Something like Gilbert Harman's framework relativism, for example, looks a lot like it is intended to be descriptive rather than revisionary (Harman & Thomson, 1996). And even when the subjectivist theory is openly revisionary such as Peter Railton's theory (Railton, 2003), the most its defender will want to admit to is that it is *tolerably* revisionary, which just means that it is not so revisionary that it is more fitting to say that we have abandoned the concept rather than revised it (Railton, 1989: 159; Miller, 2013: 185). So our Moorean arguments target theories that aspire to be less than radically revisionary, and this makes them feel a bit different, precisely because the target does not seem to be in obvious radical conflict with our common sense beliefs. Still, the form of the arguments will be identical, and, in the end, the difference between the classical Moorean arguments in the ones I employ against subjectivism does not amount to much.

2.3.1 Moorean Arguments Against Simple Subjectivism

The easiest way to see how Moorean arguments will work to show that subjectivism is radically revisionary will be to put them to use against a simple and blatantly implausible version of subjectivism. I will call this caricaturized thesis *simple subjectivism*, or SS for short. I will then show how the same kind of argument can be run no matter how simple subjectivism is modified, as long as it remains a completely subjectivist thesis. Simple subjectivism really is just the simplest form of subjectivism we can formulate, and it is supposed to capture the view that “what you think or want to do is right.” And, of course, it trivially satisfies our requirements for being a subjectivist theory.

Simple subjectivism/SS (rough description): Moral reasons are given by an agent’s desires, beliefs, preferences or goals, in other words, by an agent’s *attitudes*.

A key element of simple subjectivism is how it defines what it is to have a moral reason:

Moral reason SS: Just in case an agent A desires to, prefers to, aims to, or believes that she has a moral reason to ϕ , A has a moral reason to ϕ .

Further, it will be useful to say what it is to have a decisive moral reason on SS:

Decisive moral reason SS: Just in case A’s desire or preference to ϕ is the strongest desire or preference, or her most important goal, or if A believes that she morally ought to ϕ all things considered, then A has decisive moral reason to ϕ and ought morally to ϕ all things considered.

It is not unlikely that there are no philosophers today who believe in SS. Some philosophers attribute a view much like this to Thomas Hobbes (See for example Miller, 2013: 186; Railton, 2003: loc. 859) but, as I am not extremely familiar with his writings, I cannot say whether he actually held such a view. The point, however, is not to try to establish SS as a credible form of subjectivism—to pretend that finding an argument against SS is some sort of accomplishment would be a strawman. Rather the point is to use simple subjectivism as a base from which to build other more plausible forms of subjectivism, and to apply the Moorean arguments to them. And simple subjectivism is an ideal place to start from when we are working up towards better

subjectivisms, precisely because it is the simplest and most basic form of subjectivism we can formulate.

From here on, all the Moorean arguments I will discuss will have the form of the *reductio* version of HANDS. What we use to make Moorean arguments against simple subjectivism, then, is a *reductio* assumption, a linking premise and a Moorean premise. Now that we have a formulation of the theory, we have what is needed for the *reductio* assumption, which is just a statement of the relevant content of the theory. The thing to look for now is a Moorean premise that will form a contradiction with the *reductio* assumption given a suitable linking premise, and for that we need a belief that has what it takes to be Moorean fact.

A belief that will generate the proper contradiction is not hard at all in the case of simple subjectivism. Just consider the fact that the following claim could be true on simple subjectivism: “a person who wants to torture her neighbor could have a moral reason to do so, just because she wants to.” Intuitively, this just seems false; indeed, it seems so blatantly untrue that we could expect a statement of its contradiction to be a statement of a Moorean fact. And all it really needs to be true is simple subjectivism combined with the additional assumption that someone somewhere actually does want to torture someone without having any other moral reason to do so. It looks like we have all we need to formulate our first Moorean argument against subjectivism. Here is how the argument turns out:

- | | |
|-----------------------|---|
| TORTURE
(REDUCTIO) | <ol style="list-style-type: none"> 1. If an agent wants to ϕ and has no other moral reason to ϕ, she has a reason to ϕ just because she so wants. (From simple subjecticism) 2. Some agent somewhere wants to torture a person without having any other moral reason to torture a person. (Linking premise) 3. No agent has a moral reason to torture a person just because she so wants. (Moorean premise) <hr/> <ol style="list-style-type: none"> 4. Simple subjectivism is false. (By <i>reductio</i> from 1., 2. and 3.) |
|-----------------------|---|

I think this argument looks to be at least seemingly powerful (in the sense defined earlier). If this all seemed a bit too easy (a little bit like magical hand-waving, perhaps?), this should not worry us overmuch. In fact, because simple subjectivism is such a blatantly implausible theory, we should expect it to be easy to find arguments against it. The argument above, exploited a certain quirk with the definition of SS coupled with a fact about what we think moral reasons require.

Ordinarily we don't think that it is enough just to want to do something to get a moral reason to do it. And this is especially true in the case of torturing people, or other extreme acts. But since on SS having a moral reason to do something just is wanting to do it etc., it almost looks like we are just directly contradicting SS rather than employing a proper Moorean argument. Before we move on, therefore, I will take us through generating another Moorean argument against SS, to show a kind of Moorean argument that looks a bit more like HANDS. The strategy I'm going to employ is to look for some concrete moral fact that feels less like a principle and more like particular moral judgment. For example, I am pretty sure that Anders Behring Breivik did not have decisive moral reason to kill the people he killed on Utøya.³⁵ But if we look at the definitions of decisive moral reason on SS, it becomes clear that if it is the case that Breivik himself believed that killing those people was what he had decisive moral reason to do, he in fact had decisive moral reason to kill those people on SS. And if we believe the story that Breivik himself told of his reasons for doing what he did, it is clear that he in fact did believe that he had decisive moral reason to do what he did. And now we have all we need for our argument:

BREIVIK'S ACTION (<i>REDUCTIO</i>)	1. If an agent A believes that she morally ought to ϕ all things considered, she has decisive moral reason to ϕ . (From SS)
	2. Anders Behring Breivik believed that he morally ought to have killed the people he killed on Utøya. (Linking premise)
	3. Anders Behring Breivik did not have decisive moral reason to kill the people he killed on Utøya. (Moorean premise)
	4. Simple subjectivism is false. (By <i>reductio</i> from 1., 2. and 3.) ³⁶

This is another at least seemingly powerful Moorean argument, and it seems less dependent on exploiting the technicalities of simple subjectivism, and more in line with what we shall see is a trend with all the subjectivist theories, namely that they classify some acts that we are very sure about the moral status of wrongly. The interesting thing now, will be to see what we can do to improve upon the subjectivist theory, to try to make it less vulnerable to clashing with our strong

³⁵ Of course, I also believe that what he did was wrong.

³⁶ If you think 3. is too technical to be a Moorean premise, just consider that it is logically entailed by "It was morally wrong for Anders Behring Breivik to kill those people he killed on Utøya," and "An agent A does not have decisive moral reason to Φ , if Φ -ing is morally wrong," which is a conceptual truth.

moral beliefs in a way that enables us to generate Moorean arguments. I will now move on to the first attempt at improving simple subjectivism.

2.3.2 Improving Subjectivism: Group-Relativity

One way to minimize the possible Moorean facts that may run contrary to our subjectivism is to make the definition talk about groups of agents rather than individual agents. This will still let us stay within the bounds of subjectivism, but it might rule out Moorean arguments based on the aberrant beliefs and actions of individuals like Anders Behring Breivik. Here is an attempt to formulate a subjectivism that gives a theory that looks a lot like what we think of as relativism.³⁷

Group-relative subjectivism/GRS (rough description): Moral reasons are given by the moral beliefs of groups of agents.

Moral reason GRS: Just in case the majority of people in group G believe that there is moral reason to ϕ , people in group G has moral reason to ϕ .

Decisive moral reason GRS: For all agents A who belong to group G, just in case A is in a situation where most people in group G would think that she morally ought to ϕ all things considered, then A has decisive moral reason to ϕ and ought morally to ϕ all things considered.

Because we now make reference to what most people in the group think, and not just to individual agents, it ought to be the case that there are fewer Moorean facts available to contradict the theory. Still, it is pretty obvious that the same kind of strategy that led us to BREIVIK'S ACTION will give us an argument that looks equally good:

- | | |
|-----------------------------|--|
| ANTI-SEMITISM
(REDUCTIO) | <ol style="list-style-type: none">1. If A is in a situation where most people in group G would think that she morally ought to ϕ all things considered, then A has decisive moral reason to ϕ. (From GRS)2. People in Nazi Germany were members of a group of people where most people thought that they morally ought to discriminate against Jews. (Linking premise)3. People in Nazi Germany did not have decisive moral reason to discriminate against Jews. (Moorean premise) |
|-----------------------------|--|

³⁷ As I have already said, I think the most interesting versions of relativism are forms of subjectivism. While I do not know of any philosophers who hold exactly the view that I call group-relative subjectivism, many philosophers do espouse views that could be seen as quite close relatives of GRS. Gilbert Harman (Harman & Thomson 1996) is one such example.

As before, we have made ourselves a seemingly powerful Moorean argument, without expending much effort. Now, there are ways to reformulate group-relative subjectivism in order to evade this argument, for example by restricting the kind of group that is allowed to count. Perhaps the most radical way to do this, without leaving the bounds of subjectivism, would be to restrict the group to all of humanity. This would narrow down the list of historical examples available quite a bit, but there still seems to be some good candidates available. It is quite plausible (take this with a grain of salt as it comes from a lay historian) to say that at least at some point in history, most humans thought that it was morally mandated to grant societal privileges based on gender. But doing this is morally wrong, so this would work fine in the argument. In any case, viable Moorean facts abound if we allow ourselves access to counterfactual examples. If we do this, the amount of available Moorean facts suddenly multiply exponentially. Then we could simply imagine ourselves a group of people of the right kind as defined by our subjectivism, who collectively think or want something that we know to be wrong. For instance, we could imagine an earth where all humans wanted to and thought it was good to torture cats. Even a group-relative subjectivism that restricted the relevant group to all humans would give the answer that in this counterfactual world, torturing cats is something that humans have moral reason to do. And this would contradict the plausible Moorean fact that even if all humans liked and wanted to torture cats, torturing cats is not something we have moral reason to do.³⁸

It seems then that, no matter how we restrict or broaden the groups we are talking about, there will always be Moorean facts that in some way will contradict our definition of group-relative subjectivism. This is a highly interesting result, for it shows that there are limits to the possible success of improving subjectivism in this specific way. And this is exactly the kind of pattern that this discussion is hoping to uncover. For, if we can show that there are definite limits to how successful you can hope to be when improving subjectivism to make it more resistant to

³⁸ There is a maneuver that would make the theory immune to all Moorean arguments based on such counterfactuals. This is what is sometimes called the “rigidifying trick” (Lewis, 1989: 132), which is simply to put the operator “actual” into the definition of subjectivism. Although the maneuver is almost elegant in its simplicity, I put little stock in its actual worth. I echo Lewis’ sentiment that the rigidifying trick is just that, a trick. The addition of the actuality operator just seems objectionably ad hoc. In any case, all of the Moorean arguments I will present in this thesis will work even if we allowed a rigidifying maneuver.

Moorean arguments, at least using the type of strategy we have currently been employing, then we will be one step closer to showing that *all* subjectivist theories are vulnerable to Moorean arguments. There is still quite a few steps left before we are there, but we have at least made progress.

If you think like me now, that Moorean arguments will be available against all positions like GRS we have done what we needed to in this section. That Moorean arguments can be made against GRS and related theories is also by itself a much more interesting result than the fact that they are available against simple subjectivism. After all, GRS type views are not just a caricature that no-one believes in. If it turns out, through the discussion in 4.2, that that the Moorean arguments we employ are actually independently forceful and not just seemingly powerful, this will be an interesting result by its own right, since we then have discovered (or at least reexamined) a way of arguing against relativist theories of metaethics (or those types that are sufficiently like GRS).

Before we move on, I would like to discuss one way of restricting the group of agents whose attitudes count that might look to be able to block all Moorean arguments, but that fails in a particular way. This is the thought that we could restrict the people whose attitudes count to those people who are virtuous, for instance. Depending on how you cash out the term “virtuous,” this will indeed block all Moorean arguments. For, a natural way to interpret what we mean by those who are virtuous is something like “those who only make the right judgments, have the right attitudes, etc.”, and having this as a criterion, would make the theory immune to any sort of worry and argument formulated using the strategy we have been employing so far. (The strategy of finding something we know the moral status of, and showing that the theory will get this wrong.) With the virtuosity-requirement, the theory will get no such fact wrong. But the virtuosity-requirement has the unfortunate (for the subjectivist) consequence of making any theory that employs it objectivist. It is what I call a substantive list criterion, which, as you might remember from section 1.4, was what made theories like sensibility theory objectivist. Now, the reason why a substantive list criterion always will make a theory objectivist, is that the way the criterion ultimately works is to provide a substantive list of attitudes, for instance, that it feeds into the theory it is employed by. And any time a theory employs such a criterion, it is guaranteed that some of the claims on the theory will be grounded by the existence of this list

rather than by any subjective fact about an actual agent.³⁹ We can see this clearly if we consider the reason why one would want to include such a criterion in the first place. The only reason to include such a criterion would be to guarantee that the theory yields a certain output, regardless of the subjective facts it might take as input. So the criterion might guarantee that certain actions never come out as right. For instance, it could include a requirement that the agent has the attitude of not wanting to torture cats. And unless there are at *minimum* one case where the right output would not result given the subjective facts taken as input, and the substantive list has to do the work on its own, there just would not be any reason to include the criterion in the first place.⁴⁰

The virtuosity-requirement would indeed have to be a substantive list criterion if it was to do the job it was intended to do. It would need to take the form of a substantive list of attitudes and responses that are appropriate for a virtuous agent, and then only those agents that have attitudes and exhibit the responses that are present on the list are allowed to count. But we now see that the reference to the agents themselves, let alone their subjective facts, has become completely superfluous. It is the list that is doing the work, not the group of agents.⁴¹

2.3.3 Improving Subjectivism: Deliberation

Group-relative subjectivism is vulnerable to seemingly powerful Moorean arguments. But there are other ways of improving subjectivism, ways that tinker, not with which agents or groups of agents that are allowed to count, but with the status of the subjective facts that we take as input. In doing this, we must be careful to remember the lesson from the example of the virtuosity-

³⁹ You might object, saying “But in this case, the moral facts will still be grounded in the attitudes on this list, which are subjective facts!” This is true, but they are not, and cannot be, subjective facts about *actual agents*. Subjective facts about actual agents are always subject to change, so they are unstable in a way that entries on a substantive list cannot be, at least if the list is going to be useful. Thanks to Sebastian Watzl for pressing me on this point. My exchange with him led me to the realization that my original definition of subjectivism was too weak, and that I needed to include the actuality operator.

⁴⁰ You might think that the substantive list could do the work *in conjunction* with the other criteria, in such a way that it never grounded the resulting output all by itself, but this would be a mistake. Because the list itself carries sufficient information to specify the exact result (if it didn’t do this it couldn’t do what it is intended to do) there is just no work left for the subjective facts taken as input to do. They remain just as impotent faced with the indomitable will of the subjective list as the shaman is in front of the storm he falsely believed he called forth with his rhythmical display of superstition.

⁴¹ I think sensibility-theories are much like theories employing a virtuosity-requirement. I know their proponents must think differently, but I just can’t see how it could even come close to working otherwise. But, of course, involving a substantive list criterion while remaining non-naturalist would likely push the theory into the realm of Platonism.

requirement, which made the theory objectivist because it turned out to be a substantive list criterion. We know, therefore, that we cannot try to select the subjective facts using any sort of method where we pick and choose and really already know what we are looking for beforehand. So we need to use some sort of procedural criterion, something that keeps the output firmly grounded on the input, but still does a job of limiting the types of results that we don't want. The type of criterion we want is what I call neutral function criteria (mentioned back in 1.5), and the name expresses exactly what they need to be: a function that molds the input into a more desired shape, but that is neutral in the sense that it doesn't impose any specific output directly.

So. What sort of neutral function criteria are available? What kind of procedures can we make use of that will give the subjectivist theory a better chance of avoiding the Moorean arguments? One answer should rapidly come to mind, for there is a procedure that we all go through many times every day in order to improve on our own set of attitudes and responses: deliberation. We know that our own procedure of deliberation is going to be a neutral function criterion, because we know that when we deliberate, we never really know what the output of that deliberation will be beforehand. If we did, the deliberation itself would not have made any difference, and it would only have been a sort of play act, an empty exercise or something similar. Whenever we genuinely deliberate, we do not know what conclusion we will reach until we reach it. And here a new space for improved subjectivisms opens up: the space of *deliberative subjectivism*. The most basic version of deliberative subjectivism goes as follows:

Deliberative subjectivism/DS (rough description): Moral reasons are given by the moral beliefs of agents who have undergone a careful deliberative process.

Moral reason DS: Just in case an agent A would come to believe that she has moral reason to ϕ after thinking long and carefully about whether or not there is moral reason to ϕ , she has moral reason to ϕ .

Decisive moral reason DS: Just in case A would come to believe that she has decisive moral reason to ϕ after thinking long and carefully about all her moral reasons, A has decisive moral reason to ϕ and ought morally to ϕ all things considered.

Unfortunately for deliberative subjectivism, after all this work, the possibility for generating Moorean arguments hardly seems to have diminished at all. Actually, it looks like one of the arguments we used against simple subjectivism, namely BREIVIK'S ACTION, could straightaway apply to deliberative subjectivism as well. It seems impossible to deny that Breivik thought long

and carefully about whether it was the right thing to do to kill the people he killed on Utøya, and that he still ended up believing it was the right thing to do. The argument BREIVIK'S ACTION has lost none of its seeming power despite the added criterion of careful deliberation. Clearly, simply adding a requirement of deliberation is not enough. I can imagine that you are not surprised—after all, most of us deliberate poorly all the time, and even when we do take the time to deliberate long and carefully, we often come to the wrong conclusions. To make a truly improved subjectivism, we need something more than mere careful deliberation. Fortunately for the subjectivist, the space we opened up by introducing deliberative subjectivism is a large and intriguing one, one that will take us all the way to the very best subjectivist theories available. These theories, in all their splendor, will be the topic of the next chapter.

2.4 Summary and Preview

In this chapter, we were introduced to Moorean arguments, which we learned were classically employed to attack skeptical theories like external-world skepticism and error-theory. I talked about how skeptical theories are a subset at the extreme end of revisionary theories, and that it is also possible to employ Moorean arguments against other types of revisionary theories. I also mentioned that not all Moorean arguments are created equal, and that formulating a seemingly powerful Moorean argument against a theory does not necessarily mean that the argument has any force against that theory.

I then outlined the project of formulating increasingly better versions of subjectivism, a task that will continue in chapter 3, and of showing for each theory how it is vulnerable to Moorean arguments. The project is also an attempt to make clear the shape of the space of possible subjectivist theories, and to indicate in what directions there is hope of creating a better theory. I also identified two fundamentally different types of criteria that could be used to try to counter the possibility of formulating Moorean arguments, and argued that one of them, the substantive list criteria, is a resource unavailable to the subjectivist. The other, the neutral function criteria, is the most powerful resource the subjectivist has available.

The three subjectivist theories discussed so far have been shown to be straightforwardly vulnerable to Moorean arguments. Now, the reader might have gotten the impression that this went by a bit too quickly, especially in the case of group-relativity, which is a type of view that has at least some prominent defenders. In some sense this impression is correct, and much more

could doubtless be said in defense of group relativity.⁴² The reason for the brevity of these discussions is that I consider all of the foregoing views soft targets in this context. In other words, they are not what I consider the most interesting or plausible versions of subjectivism. The views, starting with simple subjectivism, have here been used as a dialectical device to show the way that subjectivism is systematically vulnerable to Moorean arguments, and to reveal the direction that subjectivists can go to improve their theory.

In the next chapter, I will explore two theories that exist in the space of improved subjectivist theories the discussion of deliberative subjectivism pointed us toward. Both of these theories are versions of what I call idealized deliberative subjectivism, and they are the theories defended by Bernard Williams and Peter Railton respectively. These theories are the very best that subjectivism has to offer, and yet they will face the Moorean arguments in much the same way that our three naïve subjectivisms did.

3. The Best Subjectivist Theories

Our discussion so far has taken us through three naïve versions of subjectivism, which all were vulnerable to seemingly powerful Moorean arguments. Now, the discussion moves on to what I take to be some of the best versions of subjectivist metanormative theories available: Bernard Williams's theory of internal reasons and Peter Railton's theory of an agent's non-moral good. Both of these theories are versions of what I call *idealized deliberative subjectivism*, which is a continuation and improvement of the paradigm of deliberative subjectivism. It is a striking feature of these theories, originally presented and defended in fairly different contexts, that they are so similar that, with a few not so substantial tweaks, they could have been the same theory. As per the ongoing project, I will attempt to formulate Moorean arguments against the two theories.

⁴² Though I still strongly doubt that there is any way to plausibly shore up a view like group-relativity to make it secure against Moorean arguments.

3.1 What Is Idealized Deliberative Subjectivism

We have seen that deliberative subjectivism did not fare much better than simple subjectivism against the Moorean arguments. In one sense, this should not be at all surprising, since it is completely uncontroversial that not all deliberations are equal: some are sound deliberations, while others are defective in some way, perhaps by employing bad rules of inference, by involving false beliefs, or by not accounting for all the facts. So it should be obvious that even long and careful deliberation will not produce perfect results every time. This makes a theory relying on such a criterion vulnerable to Moorean arguments. We need a further way to sharpen our neutral function operator of deliberation, so that it will get the right results more often, and preferably all of the time.

Something to notice that will help us on the way: deliberative subjectivism was different from both simple subjectivism and group-relative subjectivism in one important way: it was a *counterfactual* subjectivism. It cared not about the agents' actual desires, but about the desires that the agent would have had under certain hypothetical circumstances. With deliberative subjectivism we have moved into the realm of the merely possible, and we have formulated subjectivism in terms of the subjective facts of a certain subset of non-actual possible worlds, namely those worlds where we have gone through a careful deliberation about our relevant desires.⁴³ The obvious solution, then, is to place even stricter conditions on the counterfactual. To eliminate the possibility of deliberations being defective in any of the ways mentioned above, we simply make it a condition of the counterfactual that the deliberation is not defective in these ways. In other words, we make it a requirement that the deliberation must be ideal. The result of this chain of thought is idealized deliberative subjectivism, IDS, which is a type of view that a notable number of prominent philosophers have converged upon.⁴⁴ We could also consider the type of view that is sometimes called *Humean constructivism* a subcategory of IDS.⁴⁵ Formulated carefully, idealized deliberative subjectivism might provide some resources for avoiding the Moorean arguments. However, after I have presented what I take to be the most

⁴³ Importantly, though, the resulting moral facts on this theory are still ultimately grounded in the subjective facts about the *actual* agent, because the deliberation is directly based on those.

⁴⁴ Some notable examples include Bernard Williams (1981), David Lewis (1989), Michael Smith (1994) and Peter Railton (2003).

⁴⁵ I will not have space to discuss Humean constructivism in detail, but this type of view is really so similar in essential respects to the versions of IDS we will discuss that it makes little difference. Some notable proponents of Humean constructivism include Sharon Street (2010), James Lenman (2010) and, arguably, David Velleman (2011).

plausible versions of idealized deliberative subjectivism, I shall argue that they fail to do just that.

3.2 Williams-Style Subjectivism

In his extremely influential and important paper “Internal and External Reasons” (1981), Bernard Williams presents a compelling version of idealized deliberative subjectivism. Although this is perhaps the earliest version of a well developed IDS in the literature, it is still one of the most influential accounts. Although some of his original terminology is a bit dated, and not everything is explicitly stated, all of the resources for a very sound IDS is present in the theory presented in the paper. There, the theory is presented in the context of a discussion about what kind of reasons we can have—about whether there is such a thing as *external* reasons. Williams denies that there is such a thing as proper external reasons, and his main strategy for proving his point is by developing an account of *internal* reasons that he takes to be exhaustive of the reasons we can have. Now, Williams’s IDS is a theory of normative reasons, but not specifically moral reasons. This makes Williams’s IDS the first explicitly metanormative rather than metaethical theory we discuss. It will be interesting to note how little difference this makes to the discussion as a whole.

On to the theory. Williams’s first criterion for something being an internal reason is as follows:

Criterion 1: An internal reason statement is falsified by the absence of some appropriate element from an agent’s subjective motivational set. (Williams 1981: 101)

For Williams, an agent’s subjective motivational set (hereafter just motivational set, or *S*) is just a certain subset of the subjective facts that are true of the agent. Williams uses the word “desires” to denote all the elements in *S*, but he also says that “*S* can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent,” (1981: 105). These are precisely the types of facts that we have used to formulate our subjectivist theories, it should be clear that Williams’s criterion 1 is enough to make his theory of internal reasons subjectivist, since for any given reasons claim, criterion 1 implies that the standard of correctness must refer to at least one

subjective fact. Williams could just as well have made the same point in terms of grounding, thus:

Criterion 1': The truth of an internal-reasons claim is grounded by an appropriate element x in an agent's subjective motivational set.

Being able to think of criterion 1 in terms of grounding will be useful later.

Williams's theory of internal reasons, which I will hereafter call Williams-style subjectivism, or WSS, is like simple subjectivism in that it is a certain subset of subjective facts about the actual agent, namely the elements of the agents motivational set, that grounds claims about the reasons she has, but unlike simple subjectivism, they do not do so directly. As mentioned, there must, on Williams's theory be an appropriate connection between the reasons-statements and the agents subjective motivational set. The connection is characterized by Williams as "a sound deliberative route" (1981: 120) going from the appropriate element in the subjective motivational set to the reasons-claim. We thus get criterion 2 for something being an internal reason:

Criterion 2: There must be a sound deliberative route from element x in S to internal reasons statement r .

This makes Williams-style subjectivism look a lot like deliberative subjectivism, which says that an agent has a moral reason to ϕ just in case she would have come to believe that she has moral reason to ϕ after long and careful deliberation. But there are a couple of key differences that we can note right away. First, criterion 2 says nothing about whether or not the agent would have come to realize that there was a sound deliberative route from her motivational set to the reasons statement. This opens up the possibility that the agent could be unable to realize that she had some reason even after long and careful deliberation; it is enough that some deliberative route exists, and this route need not be available to the agent. Second, deliberative subjectivism requires that the deliberative process is long and careful, and what these requirements amount to will depend on the agent's deliberative capacities. In contrast, the requirement of soundness is to the deliberative route itself, and so is independent of the actual deliberative capacities of the agent. This requirement also allows the theory to restrict the deliberative routes that are allowed to count. Depending on how the soundness requirement is unpacked, this could mean anything from a simple requirement of logical consistency and use of valid inferences to a stronger

restriction involving a more substantive notion of rationality. Here we should recall the trap of the substantive list criterion, which is what we would end up with if we were to cash out the criterion in terms of such a substantive notion of rationality. To remain subjectivist, the soundness requirement *must* be cashed out as a neutral function criterion. That means formalistic notions like logical consistency and full instrumental rationality are the only things on the table, and to make things simpler we could collapse all of these formalistic notions into the one of full instrumental rationality. The requirement of full instrumental rationality is a criterion that serves to sharpen the procedure of deliberation itself, making sure that it is a complete process, with no mistakes made on the way. It is a criterion of *deliberative ideality*. Williams does mention at least one other requirement that falls under the criteria of deliberative ideality, namely that of a capacity for vivid imagination. As with the instrumental rationality criterion, there are only certain ways that we can interpret this if we are to stay within the realm of neutral function criteria. As the treatment of this capacity is somewhat complex, and is best illustrated using an explanatory framework I will develop when discussing Railton-style subjectivism, I will save this for then. I will therefore continue as if this has been adequately accounted for as a neutral function criterion (which it will be).

There are two remaining criteria, and both concerned with the *beliefs* that form the parameters for the deliberation:

Criterion 3: There must be no false beliefs involved in the deliberative route going from element X in *S* to R, and X must not be an element that would disappear if the agent had no false beliefs. (See Williams, 1981: 103)

and

Criterion 4: The agent must have all true beliefs that are relevant to the elements of X and to the deliberative route from X to R, and element X in *S* must not be one that would disappear had the agent had some relevant true belief that she lacks. (See Williams, 1981: 103)

So, where criterion 2 is a criterion of deliberative ideality, criteria 3 and 4 are criteria of epistemic ideality. Criteria 3 and 4 can be simplified into a single criterion of full information, requiring the agent to have no false beliefs and all relevant true beliefs. Having learned her lesson that substantive list criteria lurk behind every promising avenue, the apprehensive

subjectivist should now ask whether the criterion of epistemic ideality will turn out to be another such criterion. And as was the case with the criterion of deliberative ideality above, it depends on how we ultimately cash it out, and in this case it depends on how you interpret the phrase: “relevant beliefs.” One way you could interpret it if you are eager to have the theory provide the right answers to normative questions is to say that having all relevant true beliefs means having *all true normative beliefs*. After all, we are asking normative questions, so these seem like they are the most relevant beliefs of the bunch! And if we can say that the deliberating agent must have all true normative beliefs, then it is pretty obvious that we will end up with the right answers to all normative questions after the deliberation. Have we finally found the panacea that the subjectivist has been searching for? I hate to crush the dreams of our brave subjectivist, but this time she has bought herself a solution that gives her everything she doesn’t want. There are only two ways to make sense of the setup where the knowledge of all normative beliefs is included in the full information criterion. One: the normative beliefs that you get from the output of the deliberative process are also required as input to the process. This would mean either that the process can never get started, or that it is viciously circular. Two: all true normative beliefs are available to you before you start the process of deliberation, and they are independent of the output of the process. This means that you have to provide a completely independent account of these normative facts, which you could do, but at the expense of making Williams-style subjectivism redundant. In section 3.3.1 below I will provide a continuation of this discussion. What, then, is the proper way for the subjectivist to interpret the epistemic ideality criterion? Even if we rule out normative facts as candidates for “relevant beliefs,” it looks like the criterion flirts with objectivism, since it seems that you still might think that there is a substantive list of non-normative beliefs that are relevant to all agents, generating a substantive list criterion again. But this cannot be how it works.

To figure out what to make of the epistemic ideality criterion, we should look at the examples that Williams gives to explain criteria 3 and 4. He says: that “A may be ignorant of some fact such that if he did know it he would, in virtue of some element in *S*, be disposed to ϕ ,” (1981: 103). To put some color to this, imagine that A is an avid art collector, who has a desire to acquire a Picasso, but who does not know that there is a Picasso available at a nearby auction. In this situation our criterion of full information could kick in. Since A already has a desire to buy a Picasso, it would be relevant for her to know that there is a Picasso available at a nearby auction.

This is because the belief about the auction, together with full instrumental rationality⁴⁶ allows her to see that attending the auction is a way for her to satisfy her desire to buy a Picasso, (Smith, 1994: 157). In other words, the beliefs that are relevant in the right sense are *means-ends beliefs* that are relevant to the satisfaction of some desire that is already in the agents motivational set. And this does not violate subjectivism, because all of the added beliefs are beliefs that are called after, so to speak, by the pro attitudes in the agent's motivational set. The full information requirement is indeed intended to be a neutral function criterion, and Williams's text shows us how to interpret it as such.

So, on this theory, what does an agent have reason to do? The answer depends on the content of the particular agent's motivational set, for this is the input that we start with. In this way, Williams-style subjectivism is just like all the other subjectivisms. But, on WSS, the answer to what reasons an agent has is only *indirectly* dependent on the agents actual motivational set, for the answers to what she has reason to do is directly given by the content of the idealized version of the agents motivational set, what we might call the agent's *idealized S*. This is the set of things that she would desire to do after she has been through the procedures of idealization by criteria 2 through 3: epistemic ideality and deliberative ideality. So starting with the agents actual *S*, we can imagine her going through a process of ideal deliberation, consisting in the application of full instrumental rationality combined with vivid imagination, where she identifies all the things that she could do that would promote the satisfaction of the elements of her subjective motivational set, at the same time that she is subjected to the purging of all false beliefs and the acquiring of all means-ends beliefs relevant to her desires. According to the theory then, any of the things she could do that she identifies in this way are reasons for her. The hypothetical agent in the idealized scenario thus has a potentially vast number of things that are reasons for her to do. We should also suppose that she has a way of ordering those reasons according to priority among these reasons, and to arrive at something that she has decisive reason to do. Although Williams does not say this explicitly, the natural way to complete the theory is something like the following: The idealized agent through her idealized deliberation comes up, not only with a list of things that are reasons, but with an ordering of the *strength* of these reasons as well, so that the things that are on the top of her list will be what she has decisive reason to do. Here, then, is a statement of Williams-style idealized deliberative subjectivism:

⁴⁶ Though, in this case it looks like even a fairly minimal instrumental rationality would do.

Williams-style subjectivism/WSS (rough description): Reasons are given by the conclusions that an idealized agent with no false beliefs and all the relevant true beliefs would come to through a process of ideal deliberation, based on the agent's motivational set.

Reason WSS: A has a reason to ϕ just in case she would have wanted to ϕ after going through a process of ideal deliberation involving the criteria of both the epistemic and deliberative ideality.

Decisive reason WSS: A has decisive reason to ϕ just in case to ϕ would have been on top of the list of things she wanted to do after going through a process of ideal deliberation involving the criteria of both epistemic and deliberative ideality.

Now, to be clear, Williams-style subjectivism as formulated above may not have been *exactly* what Williams had in mind when he wrote his paper on internal reasons, but it is pretty close. It is a theory that takes subjective facts about an actual agent, in the form of the content of the agent's actual motivational set, as input, and churns them through a process of idealized deliberation to get facts about what we have normative reason to do.

3.3 A Moorean Argument From a Moral Premise

It is time to look for an argument against Williams-style subjectivism. We could proceed in the same manner that we have done with the previous theories, except now we need to look within the normative domain for a Moorean fact that will contradict the theory. Rather than doing this, however, I will use the opportunity to show how WWS connects to morality. It will turn out that we can employ a Moorean fact from the domain of morality in an argument against WWS.

WSS is a theory about reasons for action in general and not just prudential reasons.⁴⁷ Since morality is a domain concerned with reasons for action, this means that WSS is also a theory about *moral* reasons. More precisely, in this context, Williams-style subjectivism puts constraints on the moral reasons we can have. This is because WSS is a theory about what reasons in the general action-guiding sense we have, and not about what subset of those are specifically moral reasons. Williams-style subjectivism is compatible with many different theories about moral reasons, ranging from those where only a small subset of the reasons we

⁴⁷ Perhaps it is also supposed to apply to epistemic reasons although I'm not sure how this would work. In any case, it makes no difference to the present discussion, other than opening up another domain where Moorean facts could be found.

have are moral reasons, to a theory that says that all of the reasons we have are moral reasons. So working only with Williams-style subjectivism, we cannot say anything about what moral reasons we have or would have that is specific enough to formulate Moorean arguments with. However, we can formulate specific Moorean arguments using Moorean facts about things that we have moral reason to do, or not do, that Williams-style subjectivism tells us that we in fact do not have reason to do. These are the things that we think an agent would in fact have moral reason to do or not do, despite there being no element in the agent's idealized subjective motivational set which could ground (the truth of a positive statement about) this reason. For instance, I believe that *any* agent has a moral reason not to torture another being simply because she thinks torturing beings is fun. I think this claim is strong enough to be a Moorean fact. Notice also that it is a fairly weak claim in that it does not claim that the reason must be decisive. So, now we can formulate another Moorean argument against Williams-style subjectivism.

TORTURE
(REDUCTIO)

1. A has a reason to ϕ just in case she would have wanted to ϕ after going through a process of ideal deliberation involving the criteria of both the epistemic and deliberative ideality. (From WWS)
 2. It is possible for A have no desire not to torture someone simply because A thinks it would be fun even after going through such a process of ideal deliberation. (Linking premise)
 3. Any agent has a moral reason to not torture someone simply because she thinks it would be fun. (Moorean premise)
-
4. Williams-style subjectivism is false. (By *reductio* from 1., 2. and 3.)

At least to my ears, this argument is seemingly powerful. It also captures the sort of concern that causes many philosophers to be suspicious of Williams's rejection of external reasons (see for instance Scanlon's (1998: 363–374).

3.3.1 Arguing Against the Linking Premise

It is time to look at a general worry one might have about this particular type of Moorean argument. For there is something about the argument that is a bit different, at least to ANTI-SEMITISM and BREIVIK'S ACTION. This is the fact that the linking premise is a modal claim. It does, on the face of it seem very implausible to deny the truth of the modal claim, at least if we are talking about metaphysical possibility. Even on mere nomological possibility, the premise is

hard to deny since the content of A's motivational set is determined by contingent psychological facts. We must, however, consider that it is A's *idealized* motivational set that is relevant, and that the criteria of the idealization *might* just be sufficient to make the linking premises untrue. Remember that the criteria for the idealization of A's subjective motivational set are twofold, both concerning epistemic ideality and deliberative ideality. The contents of A's idealized subjective motivational set and A's actual subjective motivational set might thus be radically different. For instance, say that A actually has a strong desire to drink the contents of the glass in front of her because he believes it to be gin. In fact, however, the glass is not full of gin, but of petrol, and if A had known this, she would not have had the desire to drink its contents, (Williams 1981: 103). A's idealized subjective motivational set would in this case not contain a desire to drink the contents. We have discussed the converse case of means-ends beliefs being added before, and we also discussed the possibility of considering beliefs about normativity being part of the relevant true beliefs that the epistemic ideality criterion could add. We saw that this would, if we included all true normative beliefs, lead to circularity or redundancy. But what if we restricted the beliefs in question to moral beliefs? Then the theory would not be redundant, because it could still provide us with an informative account of what prudential reasons we have, and what normative reasons *all things considered* we have.

If we say that true moral beliefs are among the beliefs granted by the epistemic ideality criterion, and we also accept a form of internalism about moral motivation, we could perhaps make an argument that it is not in fact possible for A's idealized subjective motivational set not to contain a desire not to torture someone for the fun of it.⁴⁸ I believe this argument could have merit. The argument does however depend on moral facts being grounded in something that is independent of WSS itself, because otherwise the argument would be viciously circular (like we discussed in 3.2). Interestingly, this means that Williams-style subjectivism about reasons could be combined with some suitable (if you accept the central conclusion of this thesis: read "objectivist") theory about morality and be set to avoid arguments like TORTURE. Of course, there could still be Moorean arguments against WSS that take claims from prudential

⁴⁸ The argument would go like this: 1. The true belief that torturing someone for the fun of it is wrong is always relevant to whether or not you have a reason not to torture someone for the fun of it. 2. Believing that torturing someone for the fun of it is wrong will result in a desire not to torture someone for the fun of it (internalism). 3. It is not possible for an agent's idealized subjective motivational set to contain no desire not to torture someone for the fun of it. (From 1., 2. and the contents of Williams-style subjectivism.)

normativity as Moorean premises. Although I don't have the space to show this, I believe that such arguments are available (they will be analogous to DEPRESSION in section 3.4.2 below). This means that, to be safe from Moorean arguments, WSS would need a similar account to the one for moral beliefs above in the case of prudential beliefs. I think such an account could be had, if one wanted. The trouble with this is that WWS would be with very little real work to do, as both morality and prudential reason would have to be explained by independent substantive theories. Of course, it could still provide an account of what we have normative reason to do, all things considered, providing a way to see how moral and prudential reasons are to be weighed up against each other.⁴⁹ All in all, if WWS is to remain a form of subjectivism that could give us the kind of account of normativity or morality we are after in this discussion, it looks like it will remain vulnerable to seemingly powerful Moorean arguments.⁵⁰

3.4 Railton-Style Subjectivism

I will now look at another subjectivist theory, formulated and defended by Peter Railton (2003). Railton's theory is similar to Williams-style subjectivism in most essential matters so a short resume to point out the differences will suffice. The main difference between Railton's and Williams's subjectivism is that, where Williams's theory is an account of what (internal) *reasons* an agent has, Railton's is an account of an agent's non-moral good. Railton's account is, then, an account of *value*, specifically of non-moral value for agents.⁵¹ Couching his account in terms of value rather than reasons allows Railton to make use of it in his particular teleological (consequentialist) account of morality, but, for the present discussion, talking of value rather than reasons makes little difference.⁵² This is doubly true since Railton thinks that for something to be

⁴⁹ I am not convinced that WWS would be a suitable candidate to do such weighing, but that is another story.

⁵⁰ There is yet another way to try to avoid the Moorean attack by arguing that the linking premise of TORTURE is false. Say there are some substantive requirements as to what any being's subjective motivational set must include in order for that being to be an agent. If these requirements included things like "a desire not to torture someone for the fun of it," the linking premise would fall. This line of thought leads us straight to a view that would look a bit like Kantian constructivism (mentioned in section 1.4). But such a view would be objectivist, for reasons similar to the fate of the previous attempt at blocking the linking premise (Kantian constructivism will end up employing a substantive list criterion).

⁵¹ As this is the second subjectivist theory we discuss that is not specifically moral, it will hopefully become clear that the issues concerning subjectivism is not limited to the moral domain, but is rather a phenomenon common to normativity in general: roughly the same flavor of Moorean arguments can be employed against both subjectivism about morality and non-moral normativity.

⁵² Railton does not think that moral reasons are internal in Williams's sense.

a person's non-moral good, it must be something that can give the person a reason (a potentially motivating one) to pursue it (Railton 2003: Chapter 2). The process for arriving at a person's non-moral good on Railton's theory is, in substance, exactly like the process of arriving at an agent's internal reasons on Williams's theory. The following is true: Assuming that what an agent has non-moral reason to do is always given by what is non-morally good for the agent, what an agent has non-moral reason to do on Railton's theory will coincide with what an agent has internal reason to do on Williams's theory.

Railton's theory can be summarized as follows:

Railton-style subjectivism/RSS: An agent A's non-moral good is what an idealized counterpart of A would want herself to want were she in A's position. The idealization consists in full information, full instrumental rationality and full capacity of vivid imagination (Railton 2003: Chs. 1 and 2).

First, we can note that the requirement of full information corresponds to criteria 3 and 4 of Williams-style subjectivism, and it is really just a simpler way of saying the same thing. Note, however, that "full information" is still qualified in the sense of not including information about facts that would make the theory viciously circular, i.e. facts about what is non-morally good for the agent. The requirements of full instrumental rationality and full capacity of vivid imagination are both covered by criterion 2 of Williams-style subjectivism, though Railton is more explicit about instrumental reason being the relevant kind of reasoning involved.⁵³

Railton does not mention anything like what Williams calls the agent's subjective motivational set in particular, so it might at first glance seem like criterion 1 of Williams-style subjectivism is not fulfilled. This would be a mistake, however, as Railton clearly intends an equivalent of criterion 1 to be true of his theory. He intends his theory to be a story that explains non-moral value for agents in terms of "human motivational systems and the situations in which

⁵³ Railton's theory also the subtly important virtue of emphasizing the advisor role of the idealized counterpart of the agent by centering the account around what the idealized counterpart of A would want the actual A want, rather than what the idealized counterpart would want for her idealized self. However, this is merely to stop the theory from giving the radically wrong results of making a person's non-moral good cater to the wants of a person with significantly different properties than the person who is good it is supposed to be. The idealized counterpart of A might, for instance, not have any wish to read books since she already knows the contents of every book in the world, but for this to count against reading books being among what is non-morally good for A seems absurd.

they find themselves” (Railton 2003: 55), and an agent’s motivational set is just an instance of a “human motivational system”.

3.4.1 *The Schema for Modifying S*

As with Williams-style subjectivism, what is of real interest is how the agent’s motivational set will change through the idealization. I can think of three basic ways in which the agent’s motivational set can change thus:

1. A desire can be added.
2. A desire can be removed.
3. The ordering of the strength of desires can be changed.

All of these three types of changes can be effected by a combination of the three axes of idealization: full information, full instrumental rationality and full capacity for vivid imagination. Now, as noted, the difference between the agent’s actual subjective motivational set and the idealized subjective motivational set can be quite significant, but if we are to stay true to criterion 1 or 1’, which is the criterion that makes the theories subjectivist, the changes need to follow one simple rule: an element in the actual subjective motivational set must be essential in the explanation of the change. Let us look a bit closer at how this would work. Think back to the Picasso example: here, what explains the new desire to attend the auction is not merely the knowledge that there is a Picasso available there, but the fact that the agent already wants a Picasso and that attending the auction is a way for the agent to fulfill this prior desire. So, the actual desire to get a Picasso grounds the new desire to attend the auction. In case of the glass full of petrol, what explains the disappearance of the desire to drink the contents of the glass might be that the desire to drink the contents of the glass is itself merely an *instrumental* desire, a desire that is explained by being a means to fulfill some other desire, in this case, a desire to drink a good drink and get pleasantly drunk, or some such. If this is the case, then the knowledge that the glass is full of petrol would enable the agent to see that drinking its contents is no longer a way to satisfy the prior desire to drink a good drink and get pleasantly drunk. Again, some other member of the agents actual subjective motivational set is essential to the explanation. Both of the examples discussed so far are cases where the conditions of full information and full instrumental rationality are responsible for the changes. In both of these cases, the way the desire

is added to or subtracted from the agent's subjective motivational set follows the following schema:

Schema for modifying S:

1. There is some element E in the agent's subjective motivational set.
2. The process of idealization reveals that some desire D stands in relation R to E.
3. If D is already an element in the agent's subjective motivational set and relation R is one of D *not supporting* E, then D is removed from the agent's subjective motivational set.
4. If D is not an element in the agent's subjective motivational set and relation R is one of D *supporting* E, then D is added to the agent's subjective motivational set.

Now, we might suppose that the desire to drink the contents of the glass of petrol is not an instance of an instrumental desire, but rather what we might call an *intrinsic* or *primary* desire. In this case, the desire really just is about drinking the contents of the glass, and since that is where it stops, knowing that the content is petrol and having full instrumental rationality will not be enough to get rid of the desire (although it would probably result in a strong desire not to drink the contents of the glass, that desire being explained by being a means to fulfill prior desires not to suffer excruciating pain and not to die). What is required here is rather the capacity for full and vivid imagination about what it would be like to drink the petrol, which would result in the desire to drink it fading away. This might, on the face of it, seem like an exception to the schema for modifying S, but it is really not. For, since we can take the agent's motivational set to be just the subset of subjective facts true about the agent that are relevant to the theory at hand,⁵⁴ we must also suppose that it contains something that amounts to a disposition to both desire certain things and not desire certain other things when imagined vividly enough, and this gives us what we need to fit the process into the schema. So, we might suppose, there is a disposition to not desire anything that, vividly imagined, will feel horrible in the agent's subjective motivational set. Idealization will reveal that the desire to drink the petrol stands in the relation of not being supported by this disposition (in the sense of being disposed to go away). The case will be parallel in the event of desires being introduced by vivid imagination. So the schema holds even in the case of the desires introduced or removed through vivid imagination (this is the explanation I promised you in 3.2).⁵⁵ A notable consequence of this schema is that no new desire

⁵⁴ This is the way that "motivational set" should be understood in the text in general.

⁵⁵ The same schema with a slight modification holds for the reordering of desires:

could enter or leave the agent's subjective motivational set without the appropriate relation to some earlier element in the set. And this is exactly the result that criteria 1 or 1' require.

3.4.2 A Moorean Argument From a Normative Premise

So far what has been said applies equally to both Williams- and Railton-style subjectivism. This means that if Moorean arguments using normative facts as their Moorean premises are available against WSS, they will be available against Railton-style subjectivism and vice versa.⁵⁶

Williams, in his essay, is fairly explicit about the fact that his theory has what people might think of as counterintuitive consequences, of the type that we might be able to exploit to create Moorean arguments. In one passage, he admits that a consequence of his theory is that an agent might not always have reason to pursue what she needs. But this almost sounds like a contradiction. It certainly is counterintuitive to suppose that an agent doesn't always have reason to pursue what she needs. Williams is fairly cavalier about this, which is probably because he thinks the cases in which an agent would not have a reason to pursue her needs are such unlikely and aberrant outliers that they do not pose a problem.

Railton, on the other hand, talks a bit as if such counterexamples are not possible. Notably, he relies on a care for our own well-being to be a part of our subjective motivational set, for if it was, then we would through idealization, come to desire the things that we need to satisfy our well-being. So Railton says: "Surely our well- or ill-being is among the things that matter to us most, and most reliably, even on reflection," (2003: 13). But while this might be true on average, it is by no means necessarily or even actually always true. People who do not care a smidge about their own well-being are at *least* easily conceivable, but I will go further and say that they probably are not all that uncommon. To find them one only has to look among the people we categorize as less than sane. Among people with severe depression or other mental illnesses there are surely those who do not care about their own well-being. And even if this is not a common phenomenon among the mentally ill, it seems to me exceedingly unlikely that no

-
1. There is some element E and desires D1 and D2 in the agent's subjective motivational set.
 2. The process of idealization reveals that desire D1 and D2 stand in relation R to E.
 3. If relation R is one of D1 *supporting* E *more strongly* than D2, then D1 is ordered as supporting E more strongly than D2.
 4. If relation R is one of D1 *supporting* E *more weakly* than D2, then D1 is ordered as supporting E more weakly than D2.

⁵⁶ Assuming that what an agent has non-moral reason to do is always given by what is non-morally good for the agent, as mentioned above.

one with the psychological quirk of not caring about her own well-being was ever born. Now, even if a person has no care for her own well-being in her actual subjective motivational set, she might still get this through idealization: a desire to maintain her own well-being might be instrumental to achieve some other end she desires. A sacrificial mother who only cares about her children might come to care for her own well-being since, to take care of her children, she needs to be in good health herself. But it is not all that hard to imagine an agent for whom care for her well-being would not arise even as an instrumental desire. A well-known symptom of depression is to lose interest in things that you care about, and we can well imagine cases of severe depression where the depressed agent's subjective motivational set would contain no element to which a care for her own well-being could latch on, so to speak. For, according to the schema for modifying S, there needs to be a suitable element in the motivational set that the desire can stand in the right relation to. Accordingly, if the agent's motivational set is impoverished enough at the outset, no amount of belief-correction, instrumental rationality or vivid imagination can get the right desires in. Notice that we cannot include the fact that one ought to care about one's well-being as a part of the facts that the idealized agent would come to know under full information, for this would be an independent normative truth, something that would make Railton-style subjectivism reliant on independent normative facts, just like the example with Williams-style subjectivism earlier. So, let us say, Aubrey has a case of severe depression where she has no real desires, except perhaps a vague desire to stare out the window and not be disturbed. In this case, the process of idealization could yield no desires for the agent to get out of the depression, nor for much else for that matter. The case of Aubrey gives us the following Moorean argument:

- | | |
|-----------------------------------|--|
| DEPRESSION
(<i>REDUCTIO</i>) | <ol style="list-style-type: none"> 1. The non-moral good of an agent A is what an idealized counterpart of A would want herself to want were she in A's position. (From the content of Railton-style subjectivism) 2. The idealized counterpart of Aubrey would not want herself to want to get out of the depression. (Linking premise) 3. It is part of Aubrey's non-moral good to get out of the depression. (Moorean premise) <hr/> <ol style="list-style-type: none"> 4. Railton-style subjectivism is false. (By <i>reductio</i> from 1., 2. and 3.) |
|-----------------------------------|--|

This argument has all the features we have been looking for in our Moorean arguments so far; the argument is at least seemingly powerful. If a theory of the non-moral good of agents has the consequence that it is not part of the non-moral good of an agent in Aubrey's position to get out of her depression, then that theory looks to be radically revisionary at best.

3.4.3 The Sanity Condition

Now, I think that there is a further condition implicit in Railton's theory that would make the theory immune to this type of Moorean argument, and probably to all relevant Moorean arguments. This is a condition analogous to what Susan Wolf has called the *sanity condition*, in the context of the debate about a certain kind of free will compatibilism (Wolf, 2003). In the context of our discussion, the sanity condition would say that the hypothetical agent whose desires we look to for determining non-moral good, must not only be fully informed, rational and possessed of vivid imagination, she must also be *sane*. For Railton-style subjectivism with the sanity condition to be complete then, we need to add another axis of idealization, namely sanity.

In the case of depressed Aubrey, we can easily see how this would solve our problem: if she were not depressed, she would have had a subjective motivational set filled with desires to be happy, to not be sick, to be in good shape to pursue her other interests and so on, more than enough to block arguments like DEPRESSION. And the condition would seem to block the possibility of this type of argument altogether, because we can stipulate that any agent whose subjective motivational set is impoverished or quirky enough to get the deviant results required for arguments like DEPRESSION count as insane. The trouble with the sanity condition, though, lies in making the condition explicit enough to be useful, while not making it so substantive that it breaks the bonds of subjectivism. (Do you see where this is going yet?)

There is a sense in which all of the subjectivist theories I have discussed so far, excluding simple subjectivism, are somewhat objective.⁵⁷ Group-relative subjectivism embodies one sense of objectivity in its requirement of some level of group consensus. Deliberative subjectivism caters to a different, albeit very weak, sense of objectivity, namely the one garnered by "thinking things through". Idealized deliberative subjectivism takes this objectivity further by demanding

⁵⁷ Of course, in one sense, even simple subjectivism is straightforwardly objective since it gives a straightforwardly objective answer to the question "what does A have moral reason to do?" As we discussed in 1.2, questions in psychology have objective answers in the same sense.

that the deliberation must fulfill a list of criteria, which we could say are intended to make the deliberation itself objectively good (in a sense). But these criteria themselves cannot only play a restricted role if the theory is to remain subjectivist in the sense that we want: it still has to be the subjective facts about the actual agent that grounds the truth of the claims about what the agent has moral reason to do etc. We have already considered the possibility of including knowledge about all moral facts to be included in the full information criterion. Another mock criterion that would obviously not do is to add very specific requirements like “modify the agent’s subjective motivational set to what it would be if the agent had not been an anti-Semite.” Such a requirement might have been instrumental in blocking certain Moorean arguments, but it is clearly not something the subjectivist has access to. For one, the requirement seems objectionably arbitrary when considered apart from the fact that we find anti-Semitism morally objectionable. This serves to highlight the real reason why this option is not available to the subjectivist: it is merely a way of building in the fact that anti-Semitism is not morally right into the theory, and many of the claims resulting from the theory would, as such, be grounded in the fact that anti-Semitism is not morally right rather than in subjective facts about the actual agent. We can see this quite clearly if we notice how the schema for modifying S would be subverted if a criterion like “not being an anti-Semite” had been added to idealized deliberative subjectivism. Crucially, the schema for modifying S starts with some element in the agent’s actual subjective motivational set, and the modification happens only as a result of a relation between this element and some other element is uncovered. In contrast the way that the subjective motivational set would be modified according to the not-an-anti-Semite criterion would have to start with some substantive definition of what being an anti-Semite would involve and then follow with a conditional like: If some element E in the agent’s subjective motivational set is partly constitutive of, or a result of the agent being an anti-Semite, then E is removed, or modified to its closest non-anti-Semite relative. In the earlier schema for modifying S, desires are added or removed from the subjective motivational set because of the relation the standing to some element already in S. In this new schema, however, elements of the agent’s subjective motivational set are modified because of the way they relate to something that is external to the subjective motivational set, namely being an anti-Semite. This will violate criterion 1 or 1’, which, again, is what makes these theories subjectivist.

The sanity condition is just like the not-an-anti-semitic condition in that it will violate criterion 1 or 1'. We can see this if we consider that the schema for modifying S would be underwritten in exactly the same way if we were to incorporate the sanity condition as if we were to incorporate the not-an-anti-Semitic condition. Being sane, or, as this implies, not being insane in any of the manifold ways one can be insane, involves a substantive restriction on the types of desires, dispositions and the like you can have and not have, just like not being an anti-Semite does. Because of this, the way that the subjective motivational set of an agent would be modified through a process of idealization involving idealization to sanity would proceed in exactly the same way as would be the case for idealization to non-anti-Semitism.

All of this has been a fairly complicated way of stating a simple point: the sanity condition, and any requirements that similarly involves a substantive restriction on the types of desires that can and cannot be a part of the agent's subjective motivational set, will make a theory fail to be subjectivist. This kind of criterion is what I all along have called a *substantive list criterion*, since it involves a list of things that should and should not be a part of the agent's subjective motivational set. Any theory of normativity or morality that involves such a substantive list criterion is an objectivist, not a subjectivist theory, because some of the normative or moral claims made on such a theory will be grounded in the elements on the list rather than in subjective facts about the actual agent. Railton-style subjectivism, if it involves a sanity condition, then, is really a form of objectivism, and if it does not, it is vulnerable to Moorean arguments like other subjectivist theories.

4. Evaluating the Arguments

Up to this point, several subjectivist theories have been discussed and shown to be vulnerable to seemingly powerful Moorean arguments in a systematic way. I will now argue that the way that the theories discussed are vulnerable to Moorean arguments is likely to be generalizable to all forms of subjectivism. I will capitalize on this fact to make an argument that attacks subjectivisms's status as tolerably revisionist theories. Then, I will start on the project of

evaluating the independent force of the Moorean arguments. After discussing and tentatively rejecting the most important objection to their soundness, I will conclude that the arguments hold at least some independent force.

4.1 The Case Against Tolerable Subjectivism

In this section I will formulate an argument against the view that I will call *tolerable subjectivism*. Tolerable subjectivism is any view that both presents itself as subjectivism and as a theory that is *tolerably revisionary*. As I said in section 2.3, a view is tolerably revisionary if it is not so revisionary as to make it more fitting to say that we have abandoned the concept, or changed the topic, rather than are revising it (Railton, 1989: 159; Miller, 2013: 185). Needless to say, being less than tolerably revisionary is a heavy burden to bear for any theory, especially one as dependent on our day-to-day practices and pre-theoretical beliefs as normativity or morality. If the argument against tolerable subjectivism is successful, this will be a heavy blow for subjectivist theories in general.

It is important to note that this argument is intended to work even if one has misgivings about the individual force of the Moorean arguments in chapters 2 and 3. To this end, the following discussion uses the operator *general doubts about Moorean arguments aside*, so when I talk about “powerful Moorean arguments” in this section what I mean is Moorean arguments that seem powerful, *general doubts about Moorean arguments aside*. To get this operator to work you just need to look at the Moorean arguments, distance yourself from the theoretical mindset for a moment, and ask yourself “do I really believe in the negation of x ?” where x is the Moorean premise. If the answer is a clear “no”, then (if the rest of the argument is sound) the argument is powerful in the required sense. So, for instance, ask yourself “do I really believe that Anders Bering Breivik had decisive moral reason to kill the people he killed on Utøya?” If the answer is “no,” the argument the premise figures in is powerful in the intended sense. In section 4.2 I will discuss how the Moorean arguments fare outside this operator, but that is largely an independent issue. In theory, you could accept the conclusion to the argument against tolerable subjectivism without accepting that the Moorean arguments have independent force, and vice versa.

4.1.1 *The Inductive Case against Subjectivism*

In chapters 2 and 3 I formulated and discussed several subjectivist theories, ranging from the blatantly implausible simple subjectivism to the often sophisticated and widely regarded variants of idealized deliberative subjectivism. The upshot of each discussion has been either that the theory is vulnerable to powerful Moorean arguments (and thus are shown to be *at least* revisionary), or that it involves a criterion that makes it objectivist. I will say more about what the vulnerability to Moorean arguments show below, but for now, I will assume that being vulnerable to powerful Moorean arguments is enough to refute the theory.

Even if we assume this, the specific nature of the Moorean arguments and the vast space of possible subjectivist theories ensures that the discussion has not yet given any conclusive refutation of subjectivism. After all, there could still be a form of subjectivism, one that I have not discussed, that simply is not vulnerable to any Moorean arguments. The discussion does show (continuing with the assumption) that the theories I have discussed are either refuted or not subjectivism, which is a significant result in itself considering that the variants of idealized deliberative subjectivism I have discussed are prominent and highly regarded theories. Still, I think we can get a significantly larger result than this.

What makes subjectivist theories vulnerable to Moorean arguments is that they all seem to be *extensionally inadequate*, that is, they do not provide the right answers to what moral or normative claims are true and not. For instance, we expect and want the claim “it is part of Aubrey’s non-moral good to get out of her depression” to be true, but on Railton-style subjectivism (without the sanity condition) it comes out as false. I will try to show that there is a significant pattern to the way that the subjectivisms fail to save themselves from extensionally inadequacy and thereby the Moorean arguments. The key elements in this pattern is the tension between neutral function criteria and substantive list criteria. I will now give a more extensive account of these two types of criteria than I have done thus far.

With idealized deliberative subjectivism, we ended up with three fairly strong criteria of idealization, namely full information, full instrumental rationality and full capacity of vivid imagination. These criteria, although strong, allow the theory to keep being subjectivist because the way that they work is to serve as a kind of bridge from elements already in the agent’s motivational set to potential new elements, as per the schema for modifying S. Another metaphorical way to say this is that they act as a kind of function that takes the agent’s subjective

motivational set as input, but that it is a kind of function that would not yield a positive or negative result if the input were zero, because the function does not add or subtract anything specific by itself. Let us call criteria that work in this way *neutral function* criteria.⁵⁸ As we saw, both Williams's and Railton's subjectivisms were vulnerable to Moorean arguments despite all their neutral function criteria. This result should not be particularly surprising, since the neutral function criteria by themselves look like they cannot guarantee any specific output and our beliefs about what normative and moral claims are true or not is often quite specific. As we saw with Railton-style subjectivism and the sanity condition, the natural next step is then to add criteria that do guarantee some specific results, namely substantive list criteria.

Continuing the function-metaphor, substantive list criteria are criteria that add or subtract specific things to the input much like a part of the function that goes "add *A* and *B* to *x*" would. So, even if the input (*x*) was 0, the output would still contain both *A* and *B*. Substantive list criteria can patch up the vulnerability to Moorean arguments, but will make the theory objectivist. If we want to formulate a good subjectivist theory we will find ourselves in a bind: barring criteria subjectivists do not have access to, the output of the theory will always be completely contingent on the input. And in all the cases we have gone through, this contingency has been enough to let us formulate powerful Moorean arguments.

Is there really reason to suppose that the neutral function operator will never be enough? Consider the way in which the subjectivist employing neutral function criteria might hope to ensure the right results is to make a function that sharpens, or focuses the results down to a specific range. In terms of a mathematical function represented by a graph, we could think that we are looking for a function that will plot data points is tight together as possible. (I apologize upfront for what is probably an embarrassing display of ignorance of basic math.) It is possible to make a function that does this fairly well for a certain range of inputs. A quadratic function, for instance, would let the data points for a certain limited range of input cluster around an area corresponding to the value for the input of zero. We could then think of the facts about of what we could have moral reason to do, say, as being located in a specific area on the graph delineated by a circle (or whatever closed shape). Any data point that falls within the circle will be a result that is in line with our moral beliefs. Any point that falls outside the circle will be represent a

⁵⁸ Perhaps there is a more suitable mathematical analogue but I am not mathematically versed enough to provide it. I am, however, aware that the metaphor is not completely precise, which is why I call it a metaphor.

result that is in conflict with our moral beliefs. Now, on the admittedly stretched mathematical metaphor, no function exists that would not generate results in conflict with our moral beliefs, simply because no possible function exists that will have values that stay within an enclosed area on the graph regardless of input. But that there is no possible *mathematical* function like this does not mean that there is no neutral function operator that *could* narrow the range of results to an area that overlaps that of our strong moral beliefs. The mathematical metaphor is only that: a metaphor. Still there is one fact that would give us reason to think that such a thing is impossible after all, and the mathematical metaphor gives us a hint: The line on the graph of the mathematical function will always eventually exit the enclosed area at the far extremes of its inputs. This is just how such functions work. And although the possible inputs, in the form of subjective facts, to a neutral function criteria does not go on to infinity on both extremes, as in the mathematical case, the inputs at the two extremes might still be guaranteed to give outputs that fall outside the circle.

We have already considered one extreme, the case of minimal input, which we saw in the argument DEPRESSION. There is good reason to think that no neutral function criteria can fix a case that has sufficiently minimal input, simply because the criteria cannot add anything by itself. So if we think that there are some normative reasons for example, that any agent always has, this will contradict the results produced by a case of minimal input. And as DEPRESSION showed, we do think there are such normative reasons. The other extreme, that of maximal input, doesn't seem to make much sense in this context. But I think the case of minimal input provides an exceptionally hard nut for the subjectivist to crack. My instinct tells me that there simply is no solution to be had. Perhaps I am wrong, but the pattern, I think is clear. Neutral function criteria seem hard pressed to provide the right answer presented with a difficult input.

I submit that there is at least a fair inductive argument against this possibility of formulating a subjectivist theory that will be immune to powerful Moorean arguments. The argument is not of the kind that has been the subject of philosophical concern since the time of Hume, where you go from many observations of instances to a judgment about a universal. If the argument were supposed to be of that form, it would be tenuous indeed. After all, we have only gone through a *very* limited number of cases. But luckily, the type of induction that I am talking about is the type that I suspect is the force behind most scientific inferences: the recognition of a pattern and an extrapolation of its continued shape. This is the type of induction that lies behind

our ability to solve the typical kinds of puzzles in IQ tests: If I start to recite the following numbers, 0, 1, 1, 2, 3, 5, 8 ..., you might make the plausible inference that the next number I say will be 13. But there is no guarantee, because you can never truly be certain that it *is* the Fibonacci sequence I am reciting, and not the Schmibonacci sequence say. Still, the inference to this conclusion (that the number would be 13) would be completely reasonable. Perhaps you would like to call this abduction rather than induction, but it makes no difference. I think the inference is valid here and that it is valid for the case of subjectivism's vulnerability to Moorean arguments. There is a systematic way that they have shown themselves to be vulnerable, namely by getting the wrong results if you give them a bad input. The neutral function criteria's failure to shore up this weakness is also systematic, and it might even be completely general. And subjectivists cannot help themselves to substantive list criteria, because it would make them objectivist. I know of no subjectivist theories that will fare better against Moorean arguments than those discussed in this thesis and none looks to be forthcoming. I suspect a common misgiving regarding subjectivist theories is exactly right, namely that it is impossible for a theory of morality or normativity where what is good or what one has reason to do is completely contingent on what an agent already wants, etc., to be extensionally adequate. That it seemed like subjectivist theories could succeed after all is a result of the sometimes-implicit assumption of a substantive list criterion, which means that the theory was really objectivism in disguise.

4.1.2 Are the Subjectivist Theories Radically Revisionary?

In section 2.2 I discussed how the possibility of formulating Moorean arguments indicated that a theory was revisionary. I also discussed how revisionary theories can be anywhere from mildly revisionary to radically revisionary. That a theory is radically revisionary also means that it is not tolerably revisionary. In the previous section, I argued that there is good reason to suppose that Moorean arguments are available against any subjectivist theory. If I can show that Moorean arguments not only indicate some revisionism, but indicate radical revisionism, I will have shown that there is good reason to suppose that all subjectivist theories are radically revisionary, and hence that tolerable subjectivism is wrong.

As noted, the degree to which a theory is revisionary depends on the following factors: the number of beliefs that have to be altered, and the strength and centrality of these beliefs. Now, whenever we can formulate a Moorean argument, the belief underpinning the Moorean

premise is one that needs to be altered if we are to accept the theory. This is why Moorean arguments always indicate that the theory is revisionary. Remember that the strength of a belief amounts to how confident we are in the belief, and the centrality of a belief is how crucial the belief is to the domain of discourse that is being revised. Now, any good Moorean premise must be based on a belief that is extremely strong, otherwise the argument would just not be compelling. That is why the premise is “I have hands” rather than “I have two pairs of shoes back home.” Even though you might have confidence in the latter, it is nowhere near as strong as the confidence you have in the former. Likewise, “any agent has moral reason not to torture someone for fun” is, for me at least, in a whole other ballpark than something like, “any agent has moral reason not to display an undue amount of confidence when giving directions one is quite unsure about to people on the street.” And, in fact, whether or not you can use the belief as a basis for a good Moorean argument seems to me like a good indicator of strong your belief is, since Moorean arguments need their premises to be based on extremely strong beliefs. If we now introduce an operator to eliminate extraneous intuitions on the result we have a kind of test for whether or not a belief is strong: *General doubts about Moorean arguments aside* if you find a Moorean argument in which your belief figures as the Moorean premise compelling, that belief is *strong*. Of course, this test is not intended to be an especially interesting phenomenon outside of our current context. In our context, however, the test shows that for all of the subjectivist theories I have formulated Moorean arguments against, there is at least one extremely strong belief that they require to be revised. Of course, this assumes that you, like me, find the Moorean arguments compelling, *general doubts about such arguments aside*, but I hope this is a reasonable assumption. (If you find it hard, try setting aside the distanced theoretical mindset a bit and look at the arguments again.)

That even a single extremely strong belief will have to be revised *might* be enough to label a theory radically revisionary, but we would ideally want an even stronger case. This we would get if it turned out that the strong beliefs that would have to be revised also turned out to be central beliefs. Fortunately enough, I think that there is at least a fair case to be made in favor of this.

What does it take for a belief to be central to a domain of discourse? As I hinted at in section 2.2 when talking about kitchen utensils, a central belief is one about which you think the following: “if you take this to be false, I am no longer sure that we are talking about the same

thing.” So, if you were talking to your friend about kitchen utensils and it became clear that she did not believe that kitchen utensils ever were used in the preparation of food, you would probably ask yourself, “are we still talking about kitchen utensils?” We can use this rough test to check the centrality of a belief to a discourse: If, when talking about some domain of discourse, you discover that the one you are talking to does not believe in a certain proposition⁵⁹ you are disposed to think that you are not really talking about the same thing, then the strength of that disposition is roughly equivalent to how central that belief is to the domain in question. An example of an unquestionably central belief in the case of morality would be something like “at least part of morality is concerned with what you ought and ought not to do, i.e. with actions.” If I discovered that the person I was discussing ethics with sincerely believed that morality had nothing to do with actions, I would be compelled to think that she and I were talking about different things. Morality that is completely unconcerned with actions seems to not be morality at all. Another example of a central belief could be what Michael Smith has called the *practicality requirement* (See section 1.5), which is just the belief about morality that there must be some form of (potentially defeasible) necessary connection between moral judgment and motivation (Smith, 1994: 7). At any point here things might start to get controversial. The practicality requirement has been a locus of much debate, but I think, at least in the weakest sense, it is a central belief about morality. If you claim that there really is no connection between judging that something is right and being motivated to do it, I would probably wonder whether you had just misunderstood what judging something to be right means.

Let us now look at the beliefs that are the basis of our Moorean premises. Here is a list of the sentences used to form the Moorean premises in the main Moorean arguments in this thesis:

- Anders Bering Breivik did not have decisive moral reason to kill the people he killed on Utøya.
- People in Nazi Germany did not have decisive moral reason to discriminate against Jews.
- Any agent has a moral reason to not torture someone simply because she thinks it would be fun.
- It is part of Aubrey’s non-moral good to get out of her depression.

On the face of it, these sentences look a bit different to the sentences above. And, of course, the difference is that while the sentences above expressed claims *about* morality and normativity, ones on the list express something more like moral or normative judgments. Another way to say

⁵⁹ Or whatever equivalent, if you are skeptical of propositions.

this is that the former are metaethical or metanormative second order claims but the latter are ethical or normative first order claims. The question then becomes one about whether or not first order claims can also be central. There is perhaps reason to think that first order claims are less likely to be as central as second order claims can be. If we think about the kitchen utensils claim above as a second order claim, we might think that a first order kitchen utensils claim might be something like “a sushi knife is the best tool for cutting fish.” And, if we put it through our test from above, we clearly see that this is not a central claim. Even if we take a presumably stronger claim, like “sometimes kitchen utensils need to be cleaned” it looks like it doesn’t take us as far into centrality as the “kitchen utensils are used in the preparation of food”-claim. I do, however think that it takes us some of the way. So, how does for instance “any agent has a moral reason to not torture someone simply because she thinks it would be fun” do on the centrality test? I think it actually scores fairly high, although it might not score as well as “morality is connected to action” or the practicality requirement. If you met someone who sincerely thought that there was no moral reason not to torture someone for fun, I think you would be entitled to a great deal of skepticism about that persons understanding of morality. Indeed, I think one of the best ways to interpret such a person would be to assume that she does not mean exactly the same thing as you do by “morality”. I leave it as an exercise to the reader to run other sentences through the centrality-test, but to me the conclusion seems quite clear: all the claims look like they are central to a certain degree.

Of course, we need a bit more than simply the result that all the claims from the Moorean arguments I happened to construct are central. We need to say that all such Moorean arguments would give that result if we are to generalize beyond the subjectivisms discussed. Fortunately, it looks like this is easily done. I made no particular effort, when constructing the arguments, to pick out Moorean premises based on central claims. I just needed the claims to be strong enough to do the Moorean work. So it looks a lot like it is the case that the strength required to do the Moorean work comes together with at least a certain degree of centrality. This is not a particularly surprising result, as what is required for claims to work as Moorean premises is claims that give you the feeling that “you must be crazy if you don’t believe this.” Therefore, I think we can conclude with some confidence that claims that figure as Moorean premises in powerful Moorean arguments express beliefs that are central to the relevant domain of discourse.

Having established that to reject the Moorean premise of a powerful Moorean argument is tantamount to rejecting an extremely strong and, to a certain degree, central belief, we can now conclude that the possibility of formulating powerful Moorean arguments not only shows that the theory is revisionary, but that it is radically so.

4.1.3 *The Argument Against Tolerable Subjectivism*

We now have everything we need for our argument against tolerable subjectivism:

AGAINST TOLERABLE SUBJECTIVISM	P1:	There is good inductive reason to think that all subjectivist theories are and will be vulnerable to powerful Moorean arguments. (Section 4.1.1)
	P2:	If a theory is vulnerable to powerful Moorean arguments, it is radically revisionary. (Section 4.1.2)
	P3:	If a theory is radically revisionary, it is not tolerably revisionary.
	C:	There is good inductive reason to think that tolerable subjectivism is incoherent.

This argument shows that we have good reason to think that most of the proponents of subjectivism are wrong about their own theories, for they think that their theory is a tolerable subjectivism, but there is good reason to think that tolerable subjectivism is incoherent.

4.1.4 *Should We Reject Radically Revisionary Theories of Morality and Normativity?*

I will now say a little bit more about why it would be so bad for proponents of subjectivism if tolerable subjectivism were incoherent.

People who promote subjectivist theories are often quite upfront about the fact that their theories are going to be somewhat revisionary.⁶⁰ The thought behind them is often something like “this is the best we are going to get, so if we have to jettison a few of our beliefs to get it, then so be it.” And indeed this seems to me like a completely reasonable method of theory-building. If philosophical work like metaethics has something going for it, should we not expect that it will sometimes actually produce results that require us to revise some of our pre-theoretical beliefs? So a little bit of revisionism should not faze us. This is why they favor tolerable subjectivism. But *radical* revision makes a theory *prima facie* less attractive than mild or no revision, simply because the more radical the revision is, the likelier it is that we have

⁶⁰ David Sobel makes this point in his (2009, 342). See also Lewis (1989), Railton (2003) and Williams (1981).

simply changed the subject rather than provided a theory about the phenomenon we wanted to understand. This is one reason why tolerable subjectivists find radical revisionism *intolerable*.

All of this is not to say that being radically revisionary automatically disqualifies a theory. Modern physics, for instance, is radically revisionary about a lot of things, for instance about the density of matter in everyday macroscopic objects as we saw in BRICK. But radically revisionary theories have to weigh up for their *prima facie* implausibility by having other considerations that favor them. In the case of modern physics, there is a host of scientific research in the form of empirical evidence and theory making backing the theory, as well as the spectacular success of science itself (however you wish to quantify it) to back up the believability of this research. In the case of subjectivist theories about morality and normativity, we have no overwhelming backing of evidence. What we have are some philosophical arguments whose conclusion is that we should favor subjectivism, warts and all. I submit that at least in this context, the philosophical arguments should count a lot less than the science when weighed against the implausibility-points gained from being radically revisionary. Consider that, within first-order ethics, we can even make good philosophical arguments in support of claims like “people in Nazi Germany did not have decisive moral reason to discriminate against Jews.” So theories that require us to give up on this belief face the burden of going against good philosophical arguments and conclusions of first-order ethics. This alone is a pretty heavy burden, but we can add to it further: It is a good bet that there is a significant consensus within first-order ethics (and first order philosophy of rationality) about all of the claims used as Moorean premises in the five main arguments of this thesis, since to work as Moorean premises, the claims need to have a strong sense of being uncontroversial. Radically revisionist subjectivist theories about morality or normativity must stand in opposition to this consensus. Put a bit differently: The thing that has to be revised is not just the pre-theoretical beliefs of laypeople, but also the systematized and well-argued beliefs of ethicist.

But I think that by far the strongest reason to reject radically revisionary theories about morality or normativity is that they simply do not give us what we want out of such a theory. Looking at the test we developed for checking whether a theory was radically revisionary, we can understand that a radically revisionary theory is a theory that is likely to feel like it is not describing what we thought it was supposed to describe. A theory on which it is true that the Nazis had decisive moral reason to discriminate against Jews just does not seem like something

that we can call “morality” in the ordinary sense. Perhaps we could call it “morality*”. But what we wanted an account of was morality, not morality*. So to be relevant, the subjectivist needs an additional account of why we should care about morality*, and care about in such that it makes sense for it to replace morality. And at least initially, it does seem that the fact that on morality* it is true that the Nazis had decisive moral reason to discriminate Jews is something that should count against us caring about morality* in such a way. Note that this seems to be different to what would be the case for a parallel situation with our discourse about everyday objects as in BRICK. That it is true on modern physics that a brick consists mostly of empty space does not seem to count against our caring about modern physics. One reason for this is that modern physics does not need to *replace* our discourse about everyday objects in the same way that morality* would need to replace morality to be meaningful. This is because modern physics presents a new way of understanding the world, which we can well care about and find extremely useful in its own right, and not just because it would fill the role of our ordinary discourse about everyday objects like bricks. We could very well suppose that the understanding of everyday objects given to us by modern physics is so radically different from our ordinary discourse that we should think of it as a discourse about everyday objects* rather than about everyday objects. The new discourse of everyday objects* is tremendously useful when doing the relevant sciences, quite independently of how our old understanding of everyday objects is useful to our everyday interaction with such objects. But the same cannot be said for morality and morality*, at least not as convincingly. Morality is both useful and something we care about for its own sake, but morality*, on the other hand, seems to be something that we only would care about or would find useful in so far as it would replace morality. Of course, why we should care about morality is itself an interesting question, but whatever the answer to this question might be, it doesn’t look like it can come apart from the answer to why we should care about morality*.

To sum up: being radically revisionary is a *prima facie* reason to reject a theory. This reason is defeasible, and in the case of physics, for instance, it looks like there are good reasons to accept even radically revisionary theories. In the case of theories of morality or normativity, it looks like these reasons are weaker. First, because the reasons come from philosophical arguments rather than science, and there seems to be at least significant philosophical arguments on the other side. This means that the philosophical arguments for accepting the theory need to

both overcome the prima facie reason against radically revisionary theories, but also the plausibility of philosophical arguments for the claims that need to be revised. Second, we need a further argument as to why we should care about the morality or normativity described by the radically revisionary theory, since it is so distant from the view that it revises. In the case of theories of morality and normativity, it looks like the fact that claims like “people in Nazi Germany had decisive reason to discriminate against Jews” (the negation of the claims used in the Moorean arguments) will come out as true on the new theory will count as a reason not to care about that theory. This does not seem to be the case with scientific theories, for which there are reasons to care about independently of the reasons we cared about the discourse pre-revision. Together, these considerations give us good (albeit still defeasible) reason to reject radically revisionary theories about morality and normativity. If tolerable subjectivism is out of the question, subjectivism becomes very unattractive as a theory of morality and normativity.

4.1.5 A Response from the Subjectivist

It is now about time that the subjectivist should get a word in edgewise, and a subjectivist response to the kind of concerns I have raised in this section exists in the literature. Sharon Street, in her paper “In Defense of Future Tuesday Indifference” (2009) provides a spirited defense of a strategy that in the end, amounts to sugarcoating the bullet before biting it. In the subjectivist spirit, Street thinks that “the substantive content of a given agent's reasons is a function of his or her particular, contingently given, evaluative starting points,” (2010: 11). And, she agrees that this means that if an agent has a strange and aberrant starting point, the result will be strange and aberrant.

One of her main examples is Parfit’s “man with future-Tuesday indifference.” This man is completely ordinary, except for one fact: he does not care at all about what happens to him on future Tuesdays. A consequence is that subjectivist theories will apparently provide him with no reason to avoid agony on future Tuesdays, something that seems completely absurd. Street has a long and sophisticated account of why this and similar examples should fail to move us. Extremely roughly put, she argues that our intuitions are just not well enough attuned to allow us to make reliable judgments in cases that seem so strange and arbitrary as the man with future-Tuesday indifference. Because of this, it should not bother us that the result that the man with

future Tuesday indifference has no reason to care about avoiding agony on future Tuesdays feels counterintuitive.

I am willing to grant that street makes a fairly plausible case for the examples that she is discussing, such as the future-Tuesday indifferent man although there is no denying that she is eating a few bullets. What is worse for her response, though, is that be Moorean arguments discussed in this paper truly are not relying on cases that are anything close to us aberrant and strange as future-Tuesday indifference. Perhaps the most crucial Moorean argument in the whole thesis, DEPRESSION, only relies on the possibility of the existence of a deeply depressed person. And this is, while perhaps not an example of a *normal* agent, is surely not deeply strange or aberrant in the way required for her considerations to work.

On that happy note, let us move on to the discussion of the independent force of the Moorean arguments.

4.2 How Strong Are the Moorean Arguments?

I have discussed the consequences of subjectivist theories of morality and normativity being vulnerable to powerful Moorean arguments, *general doubts about such arguments aside*. Now, I will look at the independent force of the Moorean arguments, and discuss whether we should doubt the force of these arguments in the case their application to subjectivism about morality and normativity. To rid us of the suspense: I will argue that the Moorean arguments against subjectivism are much stronger than BRICK or WITCH for interesting reasons, but that they probably are a bit weaker than HANDS. In the end, the Moorean arguments do provide independent support of the conclusion that subjectivism about morality and normativity should be rejected.

4.2.1 An Age of “Post-Moorean Modesty”?

In recent years, Moorean arguments have garnered quite a bit of attention in the fields of epistemology and metaphysics. Particularly the focus has been on Moore’s original argument that appeared in “Proof of an External World”, which is the argument I called HANDS in section 2.1.⁶¹ There has also been wave of thinkers endorsing Moorean responses to revisionism and

⁶¹ See e.g. Thomas Kelly (forthcoming) and Duncan Pritchard (2002).

anti-realism of various kinds. Kit Fine has gone so far as to say that in “in this age of post-Moorean modesty, many of us are inclined to doubt that philosophy is in possession of arguments that might genuinely serve to undermine what we ordinarily believe,” (Fine, 2001). The debate about the status of Moorean arguments in general is a bit too extensive to get into here. For current purposes, it will suffice to say that, while Fine’s vision of a post-Moorean world might seem to overestimate the force of Moorean arguments in general, some Moorean arguments, like the original HANDS, are hard to ignore in the current philosophical landscape. If you want to defend a skeptical theory about the external world, you had better have a response to HANDS. For the discussion in the rest of this chapter then, I will take HANDS as a baseline for a *forceful*⁶² Moorean argument, where being a forceful argument simply means that it is something that really counts in favor of what it is an argument for, or vice versa. So, a forceful argument for free will compatibilism would be something the free will compatibilist can use in defense of her theory, and that the free will incompatibilist has a dialectical burden to provide some sort of response to.

In section 2.2 I provided two examples of Moorean arguments that seem like they should not count as being forceful, namely WITCH and BRICK. WITCH is an argument for witch-theory, the theory that witches exist from the (purported) Moorean fact that my grandmother is a witch, and BRICK is an argument against modern physics from the Moorean fact that bricks are not made up out of mostly empty space. Rather than seeming forceful, these arguments border on seeming facetious. Accordingly, I will take these arguments as the baseline for *unforceful* Moorean arguments. This is all in keeping with one of the conclusions of section 2.2, namely that not all Moorean arguments are created equal. In the following, my task will be to see whether the arguments against subjectivism are more like HANDS or WITCH and BRICK.

4.2. *Two What Makes a Fact “Moorean”?*

What determines the forcefulness of a Moorean argument? This has been a key question in much of the debate about Moorean arguments, and it is key to this discussion as well. The most important element in the Moorean argument is the Moorean premise, and the so-called Moorean fact that it reflects, so this is the natural place to start looking for the answer. It seems obvious

⁶² Taken as distinct from “powerful” in section 4.1, which worked under the *general doubts aside*-operator.

that a poorly selected Moorean premise will result in a poor Moorean argument, but it is not at all obvious what makes something a suitable Moorean fact. Kit Fine maintains that Moorean facts are beliefs that “we ordinarily accept” (Fine, 2001). While this might be true, it is not nearly specific enough to delineate between forceful arguments like HANDS and poor arguments like BRICK. After all, the Moorean premise in brick, “bricks are not made up of mostly empty space,” looks a lot like something we would describe as something we ordinarily accept. For Fine this might be fine, since he accepts that we live in a world of post-Moorean modesty, but for our current project this will not do. Being something we ordinarily accept might well be a necessary criterion for being a Moorean fact, but it is way too inclusive to be the criterion we are looking for.

A parameter that often comes up in the discussion of Moorean facts is our confidence in the belief.⁶³ We might think that there is a confidence criterion saying something like: a Moorean fact is a belief that you are extremely confident in. And this also seems like a necessary criterion, for a good test to check if something is *not* a Moorean fact is to ask yourself whether believe strongly in it: if you don’t it is not a Moorean fact. This criterion formed a vital part of the discussion in section 4.1.2, but it still does not seem to be specific enough for our current purposes—BRICK would still come out as forceful for people not already versed in and convinced by basic modern physics. Confidence alone cannot be enough, for the confidence needs to be properly justified for the argument itself to come out as forceful. Now, Tristram McPherson notes that “[i]nsofar as one is epistemically virtuous, one’s confidence in a claim will tend to track the quality of the evidence that one has for the truth of that claim” (2009). This insight might well provide us with an adequate criterion for Moorean facts suitable for forceful Moorean arguments: Moorean facts are facts that epistemically virtuous people are extremely confident in. This might require some explanation to make sense. Here is the thought.

Because the property of being epistemically virtuous can be defined to cover all the variables we need, this criterion looks like it could be rich enough to provide what we need to delineate between WITCH, BRICK and HANDS. If we include an adequate understanding of the relevant sciences among the criteria for being epistemically virtuous, it should be obvious how the premises of both WITCH and BRICK suddenly look a lot less likely to count as Moorean facts. A person who is well adequately versed in modern physics knows that bricks are made up of

⁶³ See McPherson (2009: 6), Huemer (2005: 116) and Enoch (2011: 118).

atoms, which, if you count only the nucleus as taking up space (since electrons are point-particles), is about 99.99999999999999% empty space (or if you take into account that the nucleus itself is composed of quarks, which are also point-particles, 100% empty space). She would therefore probably not be overly confident in the belief that bricks are not made up of mostly empty space. Equally, concerning WITCH, if we laid down precise criteria for what it would take to be a witch, we would have one of the following situations. If being a witch implies having the ability to do magical things like turning people into frogs by magic, then a scientifically literate person would know that claims about the possession of such abilities have been tested and falsified time and time again. On the other hand, if being a witch does not require magical abilities, but rather only having certain beliefs and following a certain religion, i.e. being a Wiccan, then WITCH suddenly does not seem like such a bad argument, after all. HANDS on the other hand, does come out as forceful under our new criterion, as any epistemically virtuous non-amputee would believe that she had hands—there is no relevant science that would weaken this belief. We could say that the belief in the Moorean fact *resists informed change*.⁶⁴

What about our arguments against subjectivism? Take TORTURE. Is there any relevant scientific research that should make us doubt that any agent has a moral reason not to torture someone for fun? On the face of it, there does not seem to be, but this will require some further discussion. If we accept the old is/ought-gap (or Hume’s dictum as it is sometimes called), science does not directly tell us anything about what we ought to do because it only describes what *is* the case, not what *ought* to be the case, and it is impossible to derive an ought from an is. Although some interesting counterexamples to the is/ought-gap has been proposed, none of the counterexamples are thoroughly compelling and good responses are available to the one defending the gap.⁶⁵ Anyway, I think the spirit of the distinction still stands, but for our current purposes, we actually need something a bit more than a simple is/ought-gap—we need there to be a gap between what science is in the business of discovering and the normative. It is only if we have this that we can really say that science does not tell us what we have reason to do, nor what we have moral reason to do, nor what is good and what is bad. For all of the Moorean premises in the arguments against subjectivism are based on purported *facts*, although they are all normative facts, and the is/ought-gap would therefore seem to miss the mark. What we need,

⁶⁴ Thanks to Sebastian Watzl for suggesting this useful term.

⁶⁵ See e.g. Daniel Guevara (2008), Campbell Brown (2014) and Frank Hindricks (2013).

then, is a science/normativity-gap. In fact, I think there is such a gap, and I think that this gap is one of the deep and interesting things that the is/ought-gap suggests. But simply assuming a science/normativity-gap at this point would be begging the question against one of the subjectivist's main reasons for being a subjectivist, namely the hope for a reductive normative naturalism. To avoid this, we can simply make a science/normativity-gap a potentially bridgeable one. So: There is a science/normativity-gap, meaning that scientific knowledge alone does not give you knowledge about normative facts. However, it is possible that, through philosophical argument, we can discover that normative facts are identical or reducible to scientific facts, in which case, knowledge of this fact together with the scientific knowledge would give you access to knowledge about normative facts. In other words, there is a science/normativity-gap that is bridgeable by philosophy. This allows for reductive normative naturalism to be true and avoids begging the question. The upshot of this is that, because of the science/normativity-gap, the scientific literacy of an agent would not affect her confidence in her belief that any agent has a moral reason not to torture someone for fun, and TORTURE does not fail to be a forceful Moorean argument on this regard. The discussion so far has thus provided us with a way to identify a place where WHICH and BRICK fails as forceful Moorean arguments where TORTURE, and the other Moorean arguments against subjectivism, do not. However, it has not, shown that these arguments do not fail in some other regard.

There is, of course, one way that TORTURE and the rest could fail in a similar regard even on the defeasible science/normativity-gap. If being epistemically virtuous also implies being philosophically literate and that this philosophical literacy bridges the gap in such a way that the Moorean premises of TORTURE and the rest are made doubtful. Perhaps the philosophically literate person would tend to be convinced by the arguments for subjectivism, and thus be less inclined to believe anything that would contradict it. It is a bit hard to evaluate what the philosophically literate person would tend to be convinced by. It is at least much harder than the case of the scientifically literate person, considering the lack of consensus on the big questions in philosophy, but we are still faced with the question: should we not include philosophical literacy as one of the criteria for being epistemically virtuous? After all, if we are to name the great epistemic projects of humankind, philosophy should surely be on the list next to science. This brings us to an incredibly important point, which I think is indicative of the role of Moorean arguments in philosophy. David Lewis characterized a Moorean fact as "one of those things that

we know better than we know the premises of any philosophical argument to the contrary,” (Lewis, 1996: 549), and when Kit Fine mused about the age of “post-Moorean modesty,” the ones whose modesty he was talking about were philosophers, not scientists. It seems clear then, that the answer to the puzzle of BRICK and WITCH is that these fail because they attempt to attack established scientific theories or facts, rather than philosophical theories like external world skepticism, and this merely confirms the conclusion of the previous paragraph. We should probably be careful, then, if we are to add philosophical literacy to the requirements of the epistemically literate person, but this does not mean that it cannot be done. After all, basic philosophical skills do seem like a crucial thing to have for anyone who earnestly seeks the truth like an epistemically virtuous person would. And indeed, even if we do grant the epistemically virtuous agent adequate knowledge of all relevant philosophical fields, including arguments for and against subjectivism, it does not look like this would necessarily undermine her confidence in the Moorean premises of HANDS *or* TORTURE. Keep in mind that there really is no consensus on the big issues in philosophy (Chalmers & Bourget, 2014),⁶⁶ and so the general knowledge of the relevant philosophical fields should not tip the balance toward one side or the other. Moreover, despite the lack of consensus on the big issues, significant consensus on small things. If you look away from the small number of actual external world skeptics, there is almost certainly very significant consensus on the (admittedly not very philosophical) question of whether a particular human non-amputee has hands. Equally, if you look away from the fairly small number of error-theorists and convinced bite-the-bullet subjectivists (i.e. those subjectivists who accept that their theory is radically revisionist), there is almost certainly a very significant consensus on the question of whether one has reason not to torture someone for fun, particularly among the people who do ethics. So, if we take these small-issue consensus into account, the agent might even be more confident in the Moorean premises with the philosophical literacy than without. Of course, we might think that simply knowing about the existence of seemingly strong arguments for the skeptical conclusions should make the agent less confident in her beliefs, and, indeed, I think that this is true. I do not, however, think that the lessening of confidence is something that applies to the belief in the Moorean premises in particular, but rather something more akin to a global epistemic humility: someone who is well-versed in philosophy knows that for almost any claim, you can find philosophical arguments both for and against the truth of that

⁶⁶ See the PhilPapers survey for some data to support this lack of consensus at <http://philpapers.org/surveys/>.

claim. All in all, I think that the status of the Moorean premises of HANDS and TORTURE as Moorean facts remain largely unchanged through the addition of philosophical literacy to the requirements of the epistemically virtuous agent. If anything, I think the premise of TORTURE will emerge just slightly stronger, because of the fair degree of philosophical consensus about it (this is related to the point I made in 4.1.4. about the beliefs being revised not only being laypeople's, but a significant amount of philosophers' beliefs).

4.2.3 The Debunking Objection

To be a forceful Moorean argument, the argument needs to sport a proper Moorean fact as the Moorean premise, and a fact is properly Moorean only if it is strongly believed by an epistemically virtuous person, where epistemic virtuosity includes an adequate understanding of the relevant sciences, as well as an adequate understanding of philosophy in general. We have seen that both WITCH and BRICK are undermined as forceful Moorean arguments by these criteria, but that neither TORTURE, and its relatives, nor HANDS are. So far so good. But is this all there is to the selection of Moorean facts? Evidently not: an epistemically virtuous agent who happened to believe strongly that the negation of the Moorean premise of TORTURE was true would retain much of the confidence in this belief even though she had adequate knowledge of science and philosophy. In fact, the only difference between this case and the one where the agent happens to believe in the premise rather than its negation seems to be that the negation-belief goes against the small-issue consensus among philosophers discussed above. In this does not seem to be quite enough to give one the status of Moorean fact and the other not. Unfortunately, there is not all that much we can say about this. This is because the Moorean arguments are highly indexical in an important sense: they only seem to work properly from the first person perspective. The hands of HANDS need to be *my* hands—those are the hands that I know I have. Similarly, the belief that any agent has moral reason not to torture someone for fun has to be *my* belief. It does not have the same force if that merely is what *most* people, or most philosophers, believe (although how many people believe it does count to a certain degree, as criterion above). So, in a certain sense, what can count as Moorean facts are limited to what I, the wielder of the argument, already believe strongly. This does, indeed, look like a profound limit to the use of Moorean arguments, especially in the case where the Moorean fact is a normative one, because it looks like we owe an explanation as to why or how we came to have that particular strong belief. In the case of

HANDS we could very possibly tell some story about direct acquaintance: my hands are with me everywhere, I use them everyday, I see them and I feel them and I feel with them. In the case of the normative fact, we could also try to tell some story about acquaintance or about some sort of considered judgment. But the story would be much more dubious, seemingly requiring appeal to a faculty of normative perception or of normative intuition. And this opens the gate for a common criticism of Moorean arguments in the moral and normative case: the debunking-objection.

The debunking-response to Moorean arguments goes as follows: take the strong belief that is the basis for the purported Moorean fact and provide an explanation for why the person wielding the argument has that belief, an explanation that makes it clear that the reason why this person has the belief has nothing whatsoever to do with being epistemically virtuous but as a result of something that is clearly not an epistemically sound source for beliefs. So, the “I have hands”-premise could be debunked if you could clearly show that the reason why I believe that I have hands is that someone has been sitting behind me all of my life using her own hands to fool me into thinking that I have hands, picking up stuff she thought I wanted to pick up, etc. If it can be clearly demonstrated that this is in fact the case, I should clearly stop believing that I have hands and abandon that particular Moorean argument, perhaps in favor of a more podiatrically oriented one.

In the case of normative and, more specifically, moral beliefs, there is a substantial and interesting tradition of this type of debunking arguments. Nietzsche's genealogical account of Christian morality as “slave morality” could be seen as such a debunking (Nietzsche, 1998). More recently and in a more relevant context, J. L. Mackie provided a kind of debunking argument by providing a historical and evolutionary explanation for why morality seems objective (Mackie, 1977: 44–46). It is this latter type of debunking argument that presents the greatest challenge to the Moorean arguments.⁶⁷ Without further ado, here is the evolutionary version of the argument:

⁶⁷ Both Jonas Olson (2014: ch. 7.2) and Tristram McPherson (2009: 12) take these debunking arguments to be the biggest problem with Moorean arguments that take moral facts as Moorean premises, although they are only considering Moorean arguments against error theory, not subjectivism. Thanks to Conrad Bakka for pointing me in Olson's direction, inadvertently causing me to spend 60USD on his book just to get this citation. On that note, it fitting to mention that Bakka has been, aside from Sebastian Watzl and myself, the single greatest contributor to this thesis. I am extremely grateful for all of his input and I find myself happy, and consider myself fortunate, to count this error-theorist among my friends—and philosophical opponents.

EVOLUTIONARY DEBUNKING	P1:	Having strong belief in certain moral claims conferred a significant evolutionary advantage, and these strong beliefs are widespread through the population.
	P2:	If strong belief in <i>x</i> is widespread through the population, and that strong belief would confer a significant evolutionary advantage, it is likely that this belief is widespread through the population <i>because</i> of the evolutionary advantage it conferred (because it was naturally selected for).
	P3:	Conferring evolutionary advantage is not reliably truth-tracking mechanism.
	P4:	Only beliefs that we likely have <i>because</i> of some reliably truth-tracking mechanism are epistemically sound.
	C:	Strong beliefs in certain moral claims are not epistemically sound. ⁶⁸

All of the premises of this argument are extremely plausible, and if the conclusion is true, our Moorean arguments might be in trouble. After all, it does not look like an epistemically virtuous agent ought to be very confident in a belief that is potentially epistemically unsound. To make the case a bit stronger for the debunker we could add a bit of reliabilism to the requirements for being epistemically virtuous: epistemically virtuous agents must only be confident in beliefs that they have because of reliably truth-tracking features or mechanisms. With this in place, it looks like the debunking argument will condemn Moorean arguments that take moral facts as Moorean premises, while HANDS will remain unscathed.⁶⁹

4.2.4 Improving the Debunking Argument

There are a couple of fairly obvious issues with the debunking argument as I have presented it. First is that the conclusion seems to be too weak to provide what is really needed. After all, it only says that strong beliefs in *certain* moral claims are not epistemically sound, and while this alone would give us some reason to doubt our belief in moral claims in general if we did not know how to distinguish between the epistemically sound ones and the rest, a way to do precisely this seems to be available. If we suppose that we do have access to even a limited mechanism to distinguish between true and false moral claims, it seems natural to suppose that

I do not have space to discuss the debunking argument from historical concerns, but as I take this to be much weaker than the evolutionary one, but the responses I develop below could be adapted to meet the “history version” quite easily.

⁶⁸ I am heavily indebted to Martin Ravneberg in this particular section, although he would presumably disagree with my conclusions, and perhaps also with my handling of the argument. See his Master thesis for a nice exposition of the strengths of the evolutionary debunking argument.

⁶⁹ We could also run the argument using “normative claims” rather than “moral claims.”

beliefs that are to be the basis for Moorean facts need to survive the application of this mechanism. This is not just an arbitrary addition to the requirements of Moorean facthood. Any belief that is the basis for a Moorean fact is one that we expect would survive thorough reflection. Rather than think of them as based on brute intuition, we should think of them as based on what we might call considered judgments, to borrow a Rawlsian term, where considered judgments in this context are just what they sound like: judgments that survive careful consideration. If we look at our Moorean premises again we can also see we would expect belief in these kinds of things to confer any evolutionary advantage. That all agents have moral reason not to torture someone for fun is not something that one would end up thinking about or believing unless one did at least some minimal ethical reasoning. And plausibly, if we have such a thing as a mechanism tracking the truth of moral facts, ethical reasoning would make use of that. This should be enough to make the Moorean facts in the Moorean arguments against subjectivism safe from the debunking argument unless something more is added.

The next problem with the debunking argument is that the conclusion does not follow directly from the premises. P1 to P3 will only give us the following: it is likely that strong beliefs in certain moral claims are widespread through the population because of something that is not a reliably truth-tracking mechanism. But the get from P4 to the conclusion we need not just “because”, but “only because”. For there could still be a non-competing explanation for the prevalence of the strong beliefs in terms of a reliably truth-tracking mechanism. We can see this more clearly if we consider another type of belief that might possibly have conferred significant evolutionary advantage: the belief that falling from great height is dangerous. This is a belief that has such obvious survival advantages that it seems unlikely that it was not selected for at least to some degree. Yet, it seems equally unlikely that the prevalence of this belief among humans is not also due to the fact that we have reliable and easy to access ways of finding out whether falling from a great height *is* dangerous. For instance, most of us has experienced that falling from a lesser height is somewhat dangerous, we know that increasing height means a harder impact etc., and we can extrapolate from there. So, we need some further argument establishing that there is not *also* some reliable truth-tracking mechanism that can explain our strong moral beliefs. Fortunately for our debunker, it looks like such an argument is available. In her excellent paper, “A Darwinian Dilemma for Realists Theories of Value” (Street, 2006), presents an argument against what she calls the *tracking account* of evaluative judgments, which roughly is

the account that says we tend to make certain evaluative judgments because we are in possession of a mechanism that tracks the truth of these judgments. If successful, this argument will be exactly what the debunker needs, and, in fact, it will answer both of the responses we have presented so far. This is because what we needed to make the previous response work was precisely the kind of mechanism that the tracking account posits and Street argues against. Street's argument will thus, if successful, make the debunking argument bite.

Street's argument starts with the assumption that we need there to be a cogent explanatory relation between evolutionary influences and the normative beliefs we tend to have. This is a reasonable assumption, because without such a relation, most of our normative beliefs are likely to be false "in the absence of an incredible coincidence," (Street, 2006: 16). The obvious way of accounting for this relation, then, is the tracking account, where we posit that there humans possess a mechanism for forming normative beliefs that tracks the truth of normative facts and that has been selected for in by natural selection.

The first thing to notice about this account is that it puts itself forward as a scientific explanation. It offers a specific hypothesis as to how the course of natural selection proceeded and what explains the widespread presence of some evaluative judgments rather than others in the human population. In particular, it says that the presence of these judgments is explained by the fact that these judgments are true, and that the capacity to discern such truths proved advantageous for the purposes of survival and reproduction. (Street, 2006: 18)

Street argues that, as a scientific explanation, the tracking account fares poorly against a rival account of why we make certain evaluative judgments (or have certain normative beliefs) that she calls the *adaptive link account*. According to this account, the evolutionary explanation for why we have certain normative beliefs is much more direct than in the tracking account, where the thing selected for was a mechanism that tracked truth. Here the mechanism selected for is simply one that produces certain normative beliefs without regard for their truth-value. The mechanism is analogous to reflex responses like flinching.

Such mechanisms serve to link certain kinds of circumstances—the presence of a hot surface or the visit of an insect—with adaptive responses—the immediate withdrawal of one's hand or the closing of the flytrap. Judgments about reasons—and the more primitive, "proto" forms of valuing that we observe in many other animals—may be viewed, from the external standpoint of evolutionary biology, as another such mechanism. (Street, 2006: 19)

According to Street, the adaptive link account comes out as a better scientific hypothesis than the tracking account due to three major points. First, it is more parsimonious because it does not need to posit a realm of normative facts, something the tracking account does. Second, it explains the relation between the evolutionary influences and our normative beliefs much more

clearly than the tracking account. This is a point that needs some elaboration. The cogent question to ask in this regard is: how does a mechanism tracking true normative facts come for an evolutionary advantage? This is far from obvious. And, indeed, there really does not seem to be any independent evolutionary benefit to the truth of moral beliefs, even if the beliefs themselves confer evolutionary advantage. If, say, the presumably true moral belief that “you ought to help your loved ones when they are in need” turns out to confer evolutionary advantage, it looks like it would yield exactly the same evolutionary advantage if the belief turned out to be false. As Street points out, this is not the case for many types of nonnormative belief. Take the example of our belief about the danger of falling from heights from before. Here, truth really does seem to matter. It is because having the belief results in behavior that increases the chances of avoiding the danger of falling from heights that the belief gives evolutionary advantage. If the belief is false, there is no such danger, and no benefit from the avoidance behavior. Indeed, such behavior in the absence of actual danger would probably yield a net evolutionary disadvantage.

Street’s third point in favor of the adaptive link account is that it does a better job of explaining why we have the actual moral beliefs we tend to have and why some of them are stronger than others. It provides a good explanation for why we tend to believe that we have stronger obligations toward friends and family than toward strangers, for instance. Indeed, many of the moral beliefs we do have are exactly the kind of beliefs we would expect that we had if the adaptive link account had been true. In contrast, the tracking account can offer no predictions about what beliefs we would end up with other than “mostly, those normative beliefs that are true,” which will be no help at all without some independent way of determining which beliefs these are.

In sum, the tracking account fails because it is less parsimonious, less explanatorily clear and less explanatorily powerful than the adaptive link hypothesis.

4.2.5 Responding to the Debunking Argument

The debunking argument assisted by the rejection of the tracking account provides a strong case against our Moorean arguments. How can we respond? The first thing to note is that we had to do quite a bit of philosophical work to make the debunking argument stick. It is not like the case of BRICK and WITCH where all that was needed for the Moorean facts to fall were an adequate understanding of science. And if we hold true to our earlier discussion about Moorean facts and

philosophy, we should not put too much weight on particular philosophical arguments when we evaluate the status of Moorean facts. But the debunker might still insist that his case is valid. The after all the tracking account failed as a scientific hypothesis, and all we really needed in addition was the requirement of beliefs being the result of a reliable truth-tracking mechanism to be a part of epistemic virtuosity. It would therefore be good if we could provide some more direct response to the either the debunking argument or the attack on the tracking account.

Let us start with the tracking account. Although Street's assault is impressive, there are a number of things one can say against it. First, there is an objection, which Street anticipates in her paper, that the mechanism that tracks true normative facts evolved, not as a direct result of the fitness advantage it provided, but as a byproduct, a *spandrel* (Street, 2006: 33–36). To give substance to this objection we need to provide some account of what the mechanism was a byproduct of, let's call it capacity C, and then say something about the relation between capacity C and its byproduct. The obvious move is to say that capacity C just is our capacity to reason in general. This answer is no doubt on the right track, but it just isn't specific enough to be satisfactory. A more promising answer would be our capacity to grasp a priori truths, whatever this is. This capacity, if you believe in it, is certainly something that could have conferred a fitness advantage. A priori truths that it would be useful to know could include, but are not limited to, the validity of basic mathematical and logical rules of inference (at least including deduction, but perhaps induction and abduction as well), implicit or explicit knowledge of which would be integral to most types of truth-preserving reasoning, and the ability to make truth-preserving inferences about non-normative facts is surely evolutionary fitness-conferring. The relation between the capacity to grasp a priori truths about non-normative facts and the capacity to grasp truths about normative facts is pretty straightforward: either, it just is the same capacity, or, alternatively, it is a different but very similar capacity operating in much the same way. Either way it is not hard to see how the evolution of one could have brought with it the other. Although this account could use some more refinement, I think that if we believe that we have a capacity to grasp a priori truths, this is an answer that gives us what we need. But these days many people are skeptical of such a capacity, so it would be even better if we could provide a solution that did not involve this controversial element—my next response tries to do just this.

Street's article, although it is aimed at value in general and not just moral value, is mostly talking about morality as far as the arguments are concerned. And the fact is that if we try

to field the same argumentative strategy against prudential normativity, her case is not nearly as strong—the crucial point being that it actually is quite easy to see why a mechanism tracking the truth of prudential normative facts would be naturally selected for. Let us say that there really are facts about what is non-morally good for a person, and that our ordinary idea of this is approximately correct, something involving maximization of happiness, pleasure, meaning or some such. The precise notion is not terribly important right now, since all of our ordinary ideas of a persons non-moral good give about the same result. Given the existence of facts about the persons non-moral good, there would also be facts about what it is non-morally best for a person to do in a given situation. On these assumptions, it would not be at all surprising if what would be non-morally best for a person to do would most of the time roughly coincide with actions that increase her chances of surviving and reproducing. And realizing this gives us what we need to argue that a mechanism tracking the truth of prudential normative facts could have been naturally selected for. Recall that one of the main objections against the tracking account was that it just did not make any difference fitness-wise whether or not a moral belief you had were true or not. In the case of prudential normativity, this is no longer so clear. It really is because the beliefs track the truth of what is non-morally best to do that they have a good chance of approximating what is evolutionary fitness -wise the best thing to do. Another way to think of it is as follows. Imagine the best truth tracking mechanism for acquiring beliefs and organism could have, evolutionary fitness -wise. This would be a mechanism that tracked the truth of what would be best for the organism to do evolutionary fitness -wise, and then made you do that. Any mechanism that approximates the function of this ideally fitness-conferring mechanism (better than whatever action-guiding mechanism is in place already) would be selected for. We know that we do not have an ideally fitness-conferring mechanism of action guidance, although we do have something that is at least close to a mechanism for arriving at true beliefs about what would be best to do for our evolutionary fitness. And we know this because we often do things despite the knowledge that they will decrease our odds of surviving and replicating (such as voluntary sterilization). It seems to fit the data much better to say that we have something like a general, but strongly fallible, mechanism to do what will be non-morally best for us. Moreover, the part of this mechanism that fails most often seems to be the motivational part, not the truth-tracking and belief-generating part.

The plausibility of such a truth-tracking, belief-generating mechanism is itself a direct counterexample to the argument against the tracking account, but so far only for a small subset of our normative beliefs. Luckily, if we accept that there is such a mechanism, we already have what we need to defend the tracking account about the rest of our normative beliefs as well. This is because the aforementioned mechanism makes an excellent candidate for the capacity that the truth-tracking mechanism responsible for our moral beliefs etc. can be a byproduct of. For it is not at all surprising that a mechanism evolved to track the truth of some normative facts would, as a byproduct, also track the truth of other normative facts. And that is really all we need. Of course, the tracking account is still less parsimonious than the adaptive link account, but on the other points, the score has become much more even, and parsimony does not seem to be enough to carry the case alone here. Although much more could be said, I will take it that I have provided an adequate response to Street's argument against tracking for our current purposes.

Without the assistance of the argument against the tracking account, the debunking argument is on shaky ground. It can only establish that the strength of some of our normative beliefs can be explained, in part by non-truth-tracking mechanisms, and so fails in two ways. First, it does not establish that the particular beliefs required for the Moorean arguments are held for this non-epistemic reason, and second, it does not rule out the fact that there might be a more complete explanation for the strength of the beliefs that includes the appropriate reference to a truth-tracking mechanism. Although the argument establishes the possibility that the source of our normative beliefs is non-truth-tracking and as such provides some reason to doubt them, this is not so different from the way that skeptical arguments like Descartes's dream argument provide us with *some* reason to doubt our beliefs about the external world. The debunking argument, therefore, does not establish that our Moorean arguments against subjectivism are worse off than HANDS.

4.2.6 The Status of the Moorean Arguments against Subjectivism

So far in this section I have provided an account of what makes for a forceful Moorean argument in terms of what it takes to be a Moorean fact. A Moorean fact is a fact an epistemically virtuous agent believes strongly in. Among the requirements for epistemic virtuosity are adequate knowledge of relevant scientific facts and adequate philosophical understanding. These two requirements (even just the first) are enough to disqualify BRICK and WITCH as forceful Moorean

arguments. The discussion of in sections 4.2.3 through 4.2.5 revealed another important requirement, namely that the belief not be vulnerable to debunking arguments. Although such debunking arguments have been proposed against moral and normative beliefs, I concluded that they were not successful. During this discussion, it became apparent that, in order to satisfy a reliabilist requirement for epistemic virtuosity, we need humans to have access to some sort of mechanism that reliably tracks the truth of normative and moral facts. This mechanism need not be expressed through brute intuition, but is more likely to be in effect when we engage in at least minimal normative or ethical reasoning. Because of this, it makes sense to think of what indicates Moorean facts as considered judgments, not brute intuitions. This also gives us a more or less satisfying answer to the worry I raised in the first paragraph of 4.2.3: that the normative Moorean facts are things that I just happen to believe strongly in. The Moorean premise of HANDS is a case where I can tell a convincing story about why I came to believe in it, *without undermining the confidence in the belief*. The worry is that such a story is unavailable in the case of the Moorean arguments against subjectivism. We can see the debunking argument as an attempt to tell a story about why I came to believe in the Moorean premises of the arguments against subjectivism in such a way as to undermine my confidence in them. But the story that I formed my normative beliefs through the help of a truth-tracking mechanism, which I maximized my chances of activating by engaging in ethical and normative reasoning, is a story that I can tell that does not undermine my belief. And, since I rejected Street's arguments against the required truth-tracking mechanism, this story provides at least a fairly plausible response to the worry. I am fully aware that this story is not nearly as clear or immediately convincing as the story I could tell about my hands. This reflects the fact that normativity just seems to be an epistemically problematic area, and it is why I think that the Moorean arguments against subjectivism are somewhat less forceful than our paradigmatic HANDS. Still, this is a far cry from being completely unforceful, like WITCH and BRICK, and on the whole, I think the arguments are pretty convincing. In the end, the Moorean arguments against subjectivism do provide independent reason to reject subjectivism.

4.3 Summary and Preview

In this chapter, I have presented two distinct challenges to subjectivist theories. One is the inductive case for their systematic vulnerability to Moorean arguments, and the argument against

tolerable subjectivism that follows from this. Combined with concerns that make radical revisionism an especially objectionable feature in metaethical and metanormative theories, this makes for a solid case against subjectivism. The second challenge is the one that comes directly from the Moorean arguments. In section 4.2 and its subsections, I argued that the Moorean arguments against subjectivism do have independent force, despite arguments to the contrary.

In the next chapter, I will tie the whole discussion together by evaluating the status of subjectivist theories on the whole, considering the arguments for subjectivism presented in Chapter 1 in the light of our new arguments against subjectivism.

5. Subjectivism, Rejected

At the outset of this thesis, I presented what I take to be a salient and even pressing question in metaethics: is morality an objective or subjective domain? The time has come to provide an answer to that question, by evaluating the arguments presented from both sides of the field. It is time to tally the points.

5.1 In Favor of Objectivism

Although this thesis, because of my own convictions, is slanted in favor of objectivism, there has been a marked absence of positive arguments for objectivism. This is not because they are completely non-existent (see for instance Enoch, 2011: ch. 2), though they are very few in number, but rather because they tend to take about the same form as arguments against subjectivism. This is not surprising, since subjectivism and objectivism are mutually exclusive and together cover almost the entire range of the metaethical spectrum. There are only some outlier theories, like error-theory, that fails to fall under either subjectivism or objectivism. Error-theory manages this by avoiding the criteria that makes a theory objectivist or subjectivist altogether—by denying that there are any moral facts. But aside from an error-theory of some form, there is little space left for theories to fall outside the distinction. If we are loathe to accept

error-theory, therefore, any argument against subjectivism is by extension an argument for objectivism.

The main pillar of the case for objectivism is therefore the arguments against subjectivism. Still, as I argued in section 1.3, there are rather strong intuitive grounds for believing that objectivism should be the default theory.

5.2 In Favor of Subjectivism

Although there are intuitive grounds for accepting subjectivism, they are much weaker than the intuitive grounds to accept objectivism. But there are numerous arguments available that put pressure on objectivism, and some of these are quite strong. In terms of positive arguments, the position of the subjectivist is quite similar to that of the objectivist, as most of her arguments are negative and targeted at specific versions of objectivism. This is especially true for the epistemological, ontological and metaphysical challenge. If the arguments are successful, we are faced with a choice between error-theory and subjectivism, and elect subjectivism. However, the dialectical position is a bit worse for the subjectivist in this case, since the arguments only target specific versions of objectivism, and the objectivist has the option to retreat to another version.

Some of the challenges to objectivism, however, can also be directly positive arguments for subjectivism. This includes the argument from irresolvable disagreement, and certain versions of the motivational challenge. I take the best versions of these arguments, particularly the motivational one, to be the strongest consideration in favor of subjectivism.

5.3 Against Subjectivism

The strongest reasons to reject subjectivism are the arguments presented in chapter 3 and 4 of this thesis. The first is the inductive case against tolerable subjectivism. I think this case is strong, and, if successful, it greatly diminishes the plausibility of subjectivism. If tolerable subjectivism is incoherent, it makes it hard to see why we ought to care about the morality or normativity that subjectivism describes. After all it requires us to accept moral and normative conclusions that seem completely off the mark—such as that getting out of our depression, if we are severely depressed, is something we have no reason to do. This argument would, I think, cancel out any positive reason for accepting subjectivism that we might have gotten from the

argument from motivation. For this argument, from a certain point of view, is just an attempt to show that subjectivism provides an answer to why you ought to care about morality.

Next, we have the Moorean arguments, which I concluded had independent force. These attack subjectivisms individually and specifically, and provides, I think, quite strong reason to reject the theories they are wielded against. The discussion in section 4.1 and its subsections showed that such arguments are more than likely to be available against all possible subjectivist theories not just the ones presented in this thesis. Therefore, they present a quite strong reason to reject subjectivism in general.

5.4 Conclusion

All in all, I think the rejection of tolerable subjectivism combined with the independent force of the Moorean arguments will provide sufficient reason to reject subjectivism. We are then free to embrace objectivism as long as we also can avoid error-theory. But here, the Moorean arguments perform a double duty: Since all of them take normative or moral facts as their premises, they can, with the wave of a hand, be made into arguments against error-theory instead. And the same discussion that led to the conclusion that the Moorean arguments would have force against subjectivism will likely work for the case of error-theory as well. I therefore conclude as follows: We have good reason to reject subjectivism, and good reason to accept objectivism. Morality, as it turns out, is an objective domain.

Bibliography

- Audi, R., 2004. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton University Press.
- , 2011. *New Intuitionism*. Continuum International Publishing Group.
- Bagnoli, C., 2015. *Constructivism in Metaethics*. [Online]
Available at: <http://plato.stanford.edu/archives/spr2015/entries/constructivism-metaethics/>
[Accessed 6 June 2015].
- Blackburn, S., 1973. Moral Realism. In: J. Casey, ed. *Morality and Moral Reasoning*. Methuen, p. 111–129.
- , 1984. *Spreading the Word*. Oxford University Press.
- Bliss, R. & Trogon, K., 2014. *Grounding*. [Online]
Available at: <http://plato.stanford.edu/archives/win2014/entries/grounding/>
[Accessed 6 June 2015].
- Boyd, R., 1988. How to be a Moral Realist. In: G. Sayre-McCord, ed. *Essays on Moral Realism*. Cornell University Press, p. 187–228.
- Brink, D., 1989. *Moral Realism and the Foundations of Ethics*. Cambridge University Press.
- Brown, C., 2014. Minding the Is-Ought Gap. *Journal of Philosophical Logic*, 43(1), p. 53–69.
- Chalmers, D., 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- . & Bourget, D., 2014. What Do Philosophers Believe?. *Philosophical Studies* 170:465–500 (2014), Volume 170, p. 456–500.
- Colyvan, M., 2015. *Indispensability Arguments in the Philosophy of Mathematics*. [Online]
Available at: <http://plato.stanford.edu/archives/spr2015/entries/mathphil-indis/>
[Accessed 7 June 2015].
- Dancy, J., 2013. *Moral Particularism*. [Online]
Available at: <http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/>
[Accessed 9 June 2015]
- Darwall, S., Gibbard, A. & Railton, P., 1992. Toward Fin de siècle Ethics: Some Trends. *The Philosophical Review*, 101(1), p. 115–189.
- Enoch, D., 2011. *Taking Morality Seriously*. Oxford University Press.

- , 2014. Why I am an Objectivist about Ethics (And Why You Are, Too). In: R. Shafer-Landau, ed. *The Ethical Life*. Oxford University Press.
- Version cited is the draft that is available at: <http://law.huji.ac.il/upload/WOE.pdf>
- [Accessed 9 June 2015]
- Fine, K., 2001. The Question of Realism. *Philosophers' Imprint*, 1(1), p. 1–30.
- Gewirth, A., 1968. Metaethics and Moral Neutrality. *Ethics*, 78(3), p. 214–225.
- Guevara, D., 2008. Rebutting formally valid counterexamples to the Humean “is-ought” dictum. *Synthese*, Volume 164, p. 45–60.
- Harman, G., 2000. Moral Relativism Defended. In: G. Harman, ed. *Explaining Value: And Other Essays in Moral Philosophy*. Clarendon Press, p. 3–19.
- Harman, G. & Thomson, J. J., 1996. *Moral Relativism and Moral Objectivity*. Blackwell Publishing.
- Hindricks, F., 2013. Collective Acceptance and the Is-Ought Argument. *Ethical Theory and Moral Practice*, 16(3), p. 465–480.
- Huemer, M., 2005. *Ethical Intuitionism*. Palgrave Macmillan.
- Jackson, F., 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Clarendon Press.
- Kelly, T., forthcoming. Moorean Facts and Belief Revision, or Can the Skeptic Win?. *Philosophical Perspectives*.
- Kirchin, S., 2012. *Metaethics*. Palgrave Macmillan.
- Korsgaard, C. M., 1996. *The Sources of Normativity*. Cambridge University Press.
- Leiter, B., 2002. *Nietzsche on Morality*. Routledge.
- Lenman, J., 2010. Humean Constructivism in Moral Theory. *Oxford Studies in Metaethics*, Volume 5, p. 175–193.
- Lewis, D., 1989. Dispositional Theories of Value. *Proceedings of the Aristotelian Society, Supplementary Volumes*, p. 112–137.
- , 1996. Elusive Knowledge. *Australasian Journal of Philosophy*, Vol. 74, No. 4. pp. 549–567 .
- Mackie, J. L., 1977. *Ethics: Inventing Right and Wrong*. Penguin books.
- McDowell, J., 1998. *Mind, Value, and Reality*. Harvard University Press.
- McPherson, T., 2009. Moorean Arguments and Moral Revisionism. *Journal of Ethics & Social Philosophy*, 3(2).
- Miller, A., 2013. *Contemporary Mathematics: An Introduction*. 2nd ed. Polity Press.
- Moore, G. E., 1903. *Principia Ethica*. Cambridge University Press.

- , 2008. Proof of an External World. In: E. Sosa, J. Kim, J. Fantl & M. McGrath, eds. *Epistemology: An Anthology*. Blackwell Publishing, p. 26–34.
- Nietzsche, F., 1998 (1887). *On the Genealogy of Morality: A Polemic*. Maudemarie Clark and Alan J. Swensen transl. Hackett Publishing Company.
- Olson, J., 2014. *Moral Error Theory: History, Critique, Defense*. Oxford University Press.
- Papineau, D., 2009. *Naturalism*. [Online]
Available at: <http://plato.stanford.edu/archives/spr2009/entries/naturalism/>
[Accessed 7 June 2015].
- Parfit, D., 2011a. *On What Matters, vol. 1*. Oxford University Press.
- , 2011b. *On What Matters, vol. 2*. Oxford University Press.
- Platts, M., 1980. Moral Reality and the End of Desire. In: M. Platts, ed. *Reference, Truth and Reality: Essays on the Philosophy of Language*. Routledge & Kegan Paul, p. 69–82.
- Prinz, J., 2007. *The Emotional Construction of Morals*. Oxford University Press.
- Pritchard, D., 2002. Resurrecting the Moorean Response to the Skeptic. *International Journal of Philosophical Studies*, Volume 10, p. 283–307.
- Railton, P., 1989. Naturalism and Prescriptivity. *Social Philosophy and Policy*, Volume 7.
- , 2003. *Facts, Values and Norms: Essays Toward a Morality of Consequence* Cambridge University Press.
- Ross, D., 1930. *The Right and the Good*. Oxford University Press.
- Scanlon, T. M., 1998. *What We Owe Each Other*. Harvard University Press.
- , 2014. *Being Realistic about Reasons*. Oxford University Press.
- Sidgwick, H., 1907. *The Methods of Ethics*. 7th ed. Macmillan and Company.
- Smith, M., 1994. *The Moral Problem*. Blackwell Publishing.
- Sobel, D., 2009. Subjectivism and Idealization. *Ethics*, Vol. 119, No. 2, p. 336–352 .
- Street, S., 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies*, 127(1), p. 109–166.
- , 2009. In Defense of Future Tuesday Indifference. *Philosophical Issues*, Volume 19, p. 273–298.
- , 2010. What is Constructivism in Ethics and Metaethics?. *Philosophy Compass*, Volume 5, p. 363–384.

- Sturgeon, N., 1988. Moral Explanations. In: G. Sayre-McCord, ed. *Essays on Moral Realism*. Cornell University Press, p. 229–255.
- Velleman, D., 2011. *A Reasonable Relativism*. Carl G. Hempel Lectures at Princeton University.
- Wiggins, D., 1998. *Needs, Values, Truth: Essays in the Philosophy of Value*. 3rd ed. Clarendon Press.
- Williams, B., 1981. Internal and External Reasons. In: *Moral Luck*. Cambridge University Press.
- Wolf, S., 2003. Sanity and the Metaphysics of Responsibility. In: G. Watson, ed. *Free Will*. Oxford University Press, p. 372–387.